DOCUMENT RESUME

ED 325 331 SE 051 667

TITLE Interactive Genetics Tutorial Project.

INSTITUTION Wisconsin Univ., Madison. Dept. of Curriculum and Instruction.

SPONS AGENCY Department of Education, Washington, DC.

PUB DATE 88

CONTRACT G008541039

NOTE 194p.; Document contains some light and broken type especially in Appendix D.

PUB TYPE Guides - Classroom Use - Guides (For Teachers) (052)

EDRS PRICE MF01/PC08 Plus Postage.

DESCRIPTORS Biology; *College Science; Computer Assisted Instruction; *Computer Uses in Education; Experiential Learning; *Genetics; Higher Education; Learning Strategies; Problem Solving; Science Activities; Science Education; Secondary Education; *Secondary School Science; Simulation; Statistics; Student Developed Materials; Thinking Skills; *Tutorial Programs

ABSTRACT

The Interactive Genetics Tutorial (IGT) project and the Intelligent Tutoring System for the IGT project named MENDEL supplement genetics instruction in biology courses by providing students with experience in designing, conducting, and evaluating genetics experiments. The MENDEL software is designed to: (1) simulate genetics experiments that students would face in a "wet lab"; (2) give students advice on how to solve specific genetics problems; and (3) give problem-solving advice so that students would gradually build up a model of scientific inquiry. The MENDEL system described consists of a problem GENERATOR and expert problem SOLVER, several components of a TUTOR that carry out an hypothesis checking strategy, and several interface options of the GENERATOR component through which all of the other systems components interact with the student. Appended material includes: the definitions of terms used in MENDEL; a sample problem and the logic of the SOLVER; an example of an ideal justification in a tutoring system; and two research reports ("MENDEL: An Intelligent Computer Tutoring System for Genetics Problem-Solving, Conjecturing, and Understanding" and "High School Students' Problem-Solving Performance on Realistic Genetics Problems"). (KR)

****

* Reproductions supplied by EDRS are the best that can be made *

* from the original document. *

****

ED325331

Interactive Genetics Tutorial Project

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
- Points of view or opinions stated in this document do not necessarily represent official OERI position or policy

Grantee Organization:
The University of Wisconsin
Department of Curriculum and Instruction
225 North Mills Street
Madison, Wisconsin 53706

Project Director(s):
Jim Stewart
Dept. of Curric.& Instr.
The University of Wisconsin
225 North Mills Street
Madison, Wisconsin 53706
(608) 263-4638

Michael Streibel
Dept. of Curric. & Instr.
The University of Wisconsin
225 North Mills Street
Madison, Wisconsin 53706
(608) 263-4674

SE051667

Grantee Organization:
The University of Wisconsin
Department of Curriculum and Instruction
225 North Mills Street
Madison, Wisconsin 53706

Grant No.:
G008541039

Project Dates:
Starting Date: September 1, 1985
Ending Date: January 31, 1988
Number of Months: 28

Project Director(s):

Jim Stewart
Dept. of Curric. & Instr.
The University of Wisconsin
225 North Mills Street
Madison, Wisconsin 53706
(608) 263-4638

Michael J. Streibel
Dept. of Curric. & Instr.
The University of Wisconsin
225 North Mills Street
Madison, Wisconsin 53706
(608) 263-4674

Fund Program Officer(s):
Ms. JoAnne MacDonald Grason
Mr. Rusty Garth

Grant Award:

Year 1	$ 75,139.
Year 2	$ 85,425.
Total	$160,564.

SUMMARY

In the Interactive Genetics Tutorial (IGT) project, we have developed an Intelligent Tutoring System (MENDEL) to help biology students conduct transmission genetics experiments on a computer and receive advice regarding how to design such experiments, how to solve genetics problems, and how to conduct scientific inquiry. The final MENDEL software system contains: a problem GENERATOR, an expert SOLVER, a student MODELER, an hypothesis CHECKER, a problem-solving ADVISOR, and graphic interface options of the GENERATOR. These components currently only support an hypothesis-checking tutoring strategy although modifications to these components are under way so that they support other tutorial strategies.

Principal Investigator(s):

Jim Stewart
Dept. of Curric. & Instr.
The University of Wisconsin
225 North Mills Street
Madison, Wisconsin 53706
(608) 263-4638

Michael J. Streibel
Dept. of Curric. & Instr.
The University of Wisconsin
225 North Mills Street
Madison, Wisconsin 53706
(608) 263-4674

Project Reports and Products:

MENDEL Software (written in GCLISP LM and runs on an IBM AT w/5mb)
MENDEL Paper (See Appendix A):
Streibel, M.J., Stewart, J.H., Koedinger, K., Collins, A., & Jungck, J.R. (1987). MENDEL: An intelligent computer tutoring system for genetics problem-solving, conjecturing, and understanding. *Machine-Mediated Learning*. *2*(1&2). 129-159.
MENDEL Research Report #1 (See Appendix B):
Collins, A., & Stewart, J.H. (1987). A description of strategic knowledge of experts solving realistic genetics problems. Department of Curriculum and Instruction, The University of Wisconsin. Madison, Wisconsin. April.
MENDEL Research Report #2 (See Appendix C):
Slack, S.J., & Stewart, J.H. (1988). High school students' problem-solving performance on realistic genetics problems. Department of Curriculum and Instruction, The University of Wisconsin. Madison, Wisconsin. March.
MENDEL Research Report #3:
Stewart, J.H., Collins, A., & Streibel, M.J. (1987). Model-based problem solving as a basis for a philosophy of tutoring. Department of Curriculum and Instruction, The University of Wisconsin. Madison, Wisconsin. In Progress.
MENDEL Research Report #4 (See Appendix D):
Stewart. J.H., & Dale, M. (1988). High school students' understanding of chromosome/gene behavior during meiosis. Department of Curriculum and Instruction, The University of Wisconsin. Madison, Wisconsin. March.

EXECUTIVE SUMMARY

Project:

Interactive Genetics Tutorial Project

Grantee Organization:

The University of Wisconsin
Department of Curriculum and Instruction
225 North Mills Street
Madison, Wisconsin 53706

Project Director(s):

Jim Stewart
Dept. of Curric. & Instr.
The University of Wisconsin
225 North Mills Street
Madison, Wisconsin 53706
(608) 263-4638

Michael J. Streibel
Dept. of Curric. & Instr.
The University of Wisconsin
225 North Mills Street
Madison, Wisconsin 53706
(608) 263-4674

A. Project Overview

The Interactive Genetics Tutorial (IGT) project began as an attempt to incorporate research on genetics problem solving into a computer environment so that students could conduct transmission genetics experiments and receive advice from the computer on how to design such experiments, solve genetics problems, and conduct scientific inquiry. The original environment was to have included an interactive video component but this was dropped because of the large development costs and because most of the anticipated instructional benefits could be handled by computational means (e.g., computer-generated visuals). The resulting Intelligent Tutoring System for the IGT project was named MENDEL.

B. Project Purpose

The IGT project and MENDEL software supplement genetics instruction in biology courses from high school to early graduate school by providing students with experience in designing, conducting, and evaluating genetics experiments. This is significant as students rarely have the opportunity to experience realistic genetics problem solving. The MENDEL software (the GENERATOR component) was therefore designed to simulate those kinds of genetics experiments that students would face in a "wet lab". The MENDEL software was also designed to give students advice on how to solve the specific problems that they were facing. This required that the software include an expert SOLVER, capable of solving those same problems. Finally, the MENDEL software was designed to give problem-solving advice so that students would gradually build up a model of scientific inquiry. This required that the MENDEL software contain a tutor or advisor that was guided by a tutorial strategy. Because of the complexity of such a component and the lack of relevant research findings in this area, the MENDEL software only instantiated one tutorial problem-solving strategy -- that of helping students check their hypotheses. However, a general model of scientific inquiry was incorporated into MENDEL so that other tutorial strategies could eventually be included.

C. Project Background and Origins

Commissions that have examined the status of US science education have been critical and have noted the discrepancies between the problem-solving activities of students and those of scientists (National Academy of Science, 1982). Furthermore, researchers have claimed that many college students, even after successfully completing science courses, did not have an adequate understanding of appropriate concepts and problem solving procedures (Larkin et al., 1980). And finally, researchers who have examined the effectiveness of problem-solving software have found that such programs have limited success -- although real potential -- in improving the learning and performance of students (Kulik et al., 1980, 1983).

These critics and researchers have concluded that, although problem solving was a necessary component of science education, more problem solving *per se* was not sufficient. Problem solving, they concluded, must preserve and reveal the complexity of the domain that it addressed and foster understanding (Frederickson, 1984). These conclusions led us to propose a genetics experimentation environment that would remain true to the complexity of the content of genetics as well as true to the processes of problem solving and designing experiments in this domain. Our initial goal was to create a genetics experimentation environment that permitted students to design and evaluate genetics experiments and to receive advice about each stage of the process. We have succeeded in achieving the first half of our goal. However, because of the complexity of specifying all the conditions and actions for tutoring, we reformulated the second half of our original goal and focused more on helping students structure their genetics knowledge and on coaching students on their problem solving and experimentation activities. Hence, we have only instantiated an hypothesis-checking tutorial strategy in our computer environment.

D. Project Description

The current MENDEL system consists of a complete problem GENERATOR and expert problem SOLVER, several components of a TUTOR that carry out an hypothesis checking strategy (e.g., an hypothesis CHECKER, a student MODELER, and an ADVISOR) and finally several interface options of the GENERATOR component through which all of the other systems components interact with the student.

E. Project Results

One of the significant results of the project is the increase in cooperation between our department (Curriculum & Instruction) and the departments of Computer Science, Zoology and Genetics. In addition, the IGT project has had an effect on a national biology education project, BioQUEST, a consortium that is developing a university freshmen biology course around 15-18 biology simulation programs. Our link with BioQUEST will allow us to influence a national software development project that has the potential to have a positive effect on biology education. Another effect of the project has been on high school genetics education. This has occurred because of the number of high school biology teachers (over 100 in the past two years) that we have worked with in summer institutes. By using the GENERATOR we have been able to help them think differently about the teaching of genetics.

We have disseminated the results of the project via paper presentations, journal articles and technical reports (a complete listing of these dissemination activities can be found in Appendices A-D). In addition, we are disseminating the

GENERATOR through the University of Wisconsin WISCware organization. By making the GENERATOR available in this format, the philosophy that underlies the IGT project will be disseminated more widely than can the MENDEL program.

We have evaluated the component parts of the IGT, particularly the GENERATOR, SOLVER, and hypothesis entry facility for the goodness of fit to genetics. This has been done by involving geneticists as evaluators. In addition, the GENERATOR has been extensively used with both university and high school students. This work has served as a formative evaluation that has provided us with insights for revising the GENERATOR.

Because of the unanticipated length of time that it took to develop the GENERATOR, the expert SOLVER, and the hypothesis entry facility we have been unable to do a summative evaluation of the entire MENDEL system. We have however, submitted a proposal to NSF to allow us to complete work on MENDEL including a thorough evaluation of student use of it. We expect to begin this phase of the project in the Fall of 1988.

F. Project Summary and Conclusions

Some of the most important insights we have gained have been about the teaching and learning of genetics -- these now influence our thinking outside the IGT project. Foremost among these have been insights about the organization and structure of genetics that are necessary to facilitate model-based problem solving. We have found it interesting that these insights are not made explicit in university genetics instruction. A second insight has been the realization of the importance of having students solve realistic genetics problems of the type produced by the GENERATOR. By solving such problems, students engage in the 3P's of science education (Peterson & Jungck, et.al., 1988) -- problem posing, problem solving, and persuasion of peers.

A second category of insights relate to our growing understanding of the role that tutoring systems, specifically MENDEL, might have for classrooms in the immediate future. Early in the project we realized the difficulty of specifying all of the possible conditions and actions that the TUTOR might take in our unstructured problem-solving environment. Therefore we have focused more on helping students structure their genetics knowledge and on coaching students on their problem solving and experimentation activities. What all of this has meant is that we have come to be more concerned about the role of instructors in the MENDEL system. A related insight is that we now feel that it is possible to provide a great deal of "tutoring" via the tools that we provide students to use.

Another insight that we have gained that should be of value to others interested in our software is the need to spend extended periods of time with potential users of the software.

FINAL REPORT

A. Project Overview

The Interactive Genetics Tutorial (IGT) project began as an attempt to incorporate current research on genetics problem solving into a computer environment so that students could conduct transmission genetics experiments and receive intelligent advice from the computer on how to design such experiments, solve genetics problems, and conduct scientific inquiry. The original environment was to have included an interactive video component but this aspect of the project was dropped because of the large development costs (e.g., $100,000. for half an hour of interactive video) and because most of the anticipated instructional benefits could be handled as computer-generated visuals. The resulting Intelligent Tutoring System for the IGT project was called MENDEL in honor of Gregor Mendel, the originator of transmission genetics.

B. Project Purpose

The IGT project and resulting MENDEL software supplements the instruction of students in biology courses from high school to early graduate school by providing students with a laboratory experience in designing, conducting, and evaluating genetics experiments. This is significant as these students rarely have the opportunity to experience realistic genetics problem solving. The MENDEL software (the GENERATOR component) was designed to simulate those kinds of genetics experiments that students would face in a "wet lab". The MENDEL software was also designed to give students advice on how to solve the specific problems that they were facing. This required that the MENDEL system contain an expert SOLVER component capable of solving these same problems. Finally, the MENDEL software was designed to give general problem-solving advice so that students would gradually build up a model of scientific inquiry out of many problem-solving experiences. This required that the MENDEL software contain a tutor or advisor that was guided by a tutorial strategy. Because of the complexity of such a component and the lack of relevant research findings in this area, the MENDEL software only instantiated one tutorial problem-solving strategy -- that of helping students check their hypotheses. However, a general model of scientific inquiry was incorporated into MENDEL to allow other tutorial strategies to be added at a later date.

C. Project Background and Origins

Commissions that have examined the status of US science education have been critical and have noted the discrepancies between high school and university science activities and the problem-solving activities of scientists (National Academy of Science, 1982). Furthermore, researchers have claimed that many college students, even after successfully completing science courses, lacked an adequate understanding of appropriate concepts and problem solving procedures (Larkin et al., 1980). And finally, researchers who have examined the effectiveness of problem-solving software have found that such programs have limited success, although real potential, for improving student learning (Kulik et al., 1980, 1983).

These critics and researchers have concluded that, although problem solving was a necessary component of science education, more problem solving *per se* was not sufficient -- it must preserve and reveal the complexity of the domain that it addressed and foster understanding (Frederickson, 1984). These conclusions led us to propose a genetics experimentation environment that would remain true to the complexity of the content and problem-solving processes of transmission genetics.

Our initial goal was to create a genetics experimentation environment that permitted students to design and evaluate genetics experiments and to receive advice about each stage of the process. We have succeeded in achieving the first half of our goal. However, because of the complexity of specifying all the conditions and actions for tutoring we reformulated the second half of our original goal and focused more on helping students structure their genetics knowledge and on coaching students on their problem solving and experimentation activities. Hence, we have only instantiated an hypothesis-checking tutorial strategy in our computer environment. This strategy compares the student's and the SOLVER'S hypothesis on five dimensions. In doing this we have made both a pragmatic choice and a theoretical choice.

The pragmatic choice means that we have been able to create a working system which addresses some but not all of the issues involved in using computers to tutor students in genetics. The theoretical choice means that we now believe that helping students structure their genetics knowledge as they solve problems, is more fundamental than helping them learn from an expository tutoring approach. Our current research and developmental efforts are therefore focused on creating a computer environment that helps students develop models of the content domain and models of the inquiry process. A brief description of content models and inquiry models will help provide a context for the description of our project (See Figure 1).

A **content-domain** model helps one explore the world by anticipating previously unencountered data and by permitting the construction of multiple explanations for these data. A content model, therefore, is the vehicle through which one understands and learns about the world. In transmission genetics, a content model includes symbolic representations for: objects (such as chromosomes), states (such as heterozygous), and processes (such as meiosis) which are responsible for the objects' changing states. An **inquiry model** directs one's problem-posing, and design of experiments to determine which aspects of the content model can account for the observed data. Problem solving in genetics therefore combines both content and inquiry models to achieve a justified interpretation of the data.

D. Project Description

The current MENDEL system consists of a completed problem GENERATOR and expert problem SOLVER, several components of a TUTOR that carry out an hypothesis checking strategy (e.g., an hypothesis CHECKER, a student MODELER, and an ADVISOR), and finally several interface options of the GENERATOR component through which all of the other systems components interact with the student. Figure 2 summarizes these components.

```
                      PROBLEM-SOLVING IN TRANSMISSION GENETICS
                                 /               \
                                /                 \
                               /                   \
                        Content Models       Inquiry Models
                             /                        \
                            /                          \
MENDEL-    |               /                            \
provided   |  Models of Genetics Objects,       Model of Scientific
models     |     Processes, and States                 Inquiry
                        /\                                /\
                       /  \                              /  \
                      /    \                            /    \
MENDEL-    |    Expression Charts                      /      \
provided   |     Punnett Squares                   Strategy Trees
notational|     Pedigree Diagrams               Problem-Solving Rules
systems    |  Chromosome-Pair Graphs           Justification Heuristics
                 /               \                 /              \
                /                 \               /                \
               /                   \             /                  \
              /            --Solving Genetics Problems--             \
   ------------------      | - performing crosses       |      ------------------
   |  Knowledge of  |      | - redescribing phenotype   |      |   Justified    |
   |  Meiosis and   |      |      data                  |      | Interpretation |
   |    Genetics    |      | - generating & testing     |      |     of the     |
   | Acquired from  |<---->|      hypotheses            |<---->|   Phenotypic   |
   |    Lectures,   |      | - consistency checking     |      |    Data in     |
   | Readings, and  |      | - alternate hypothesis     |      |  Model-Based   |
   |  Instructors   |      |      disconfirmation       |      |     Terms      |
   ------------------      ------------------------------      ------------------

  Student's knowledge      Student's activities that are       Student's Inter-
  that is brought to       carried out within the MENDEL       pretations of
  the MENDEL system        system                              MENDEL's "experi-
                                                               mental data"
```

Figure 1

The Problem GENERATOR Component of MENDEL. The GENERATOR program contains a problem-customization section and a problem-solving environment. It is an extension of the GENETICS CONSTRUCTION KIT (Jungck & Calley, 1986). The problem-customization section permits a student or instructor to construct: a class of organisms with **trait** and **variation** names; and, the genetic parameters for that class of organisms. The genetics parameters consist of: the actual number of traits, variations, and range of progeny that could result from any cross; and, the **inheritance patterns** (simple dominance, co-dominance, multiple alleles, and gene interaction) and the **modifiers** of these inheritance patterns (sex linkage, lethality, penetrance, pleiotropy, autosomal linkage, and interference).

Once a class of problems has been created in the customization section, students use the GENERATOR's problem-solving environment without knowing the specific customization parameters. Problems are presented to the student on the computer screen as a phenotypic description of an initial population of organisms (See Figure 3).

The students can then produce offspring and perform statistical tests on this offspring data until they are satisfied with their inferences about the inheritance patterns and modifiers that could account for the offspring phenotype patterns. Thus, decisions such as whether enough data has been collected, or what the results of statistical tests mean, must be made by students. The significant feature of this type of software is that it allows students to assume the responsibility for the design and interpretation of their own experiments. To do this students have to understand the meaning the genetics terms involved (i.e., objects, states, and processes) and know how to use the associated concepts to solve genetics problems. These genetics terms also have to be represented internally in a form that the LISP language can manipulate. Hence, the GENERATOR uses definitions and symbolic representations for these objects, processes, and states. Additional information on these definitions can be found in Appendix E. Once these constructs are created in the LISP language, they are combined in the GENERATOR program into "frames" (or lists) that represent larger constructs such as: **Population**; **Chromosome**; **Locus**; **Trait**; **Offspring** and; **Sex Class**. Extended definitions of these constructs is provided in Appendix F.

The GENERATOR program then carries out several high-level functions (**CUSTOMIZATION**, **MAKING INITIAL POPULATION**, and **CROSSing**) which allow the student to interact with a particular problem. These functions are also described in Appendix F.

The Expert Problem SOLVER Component of MENDEL. The SOLVER program incorporates rules for solving genetics problems and strives to account for the same data as confronts the student. It therefore tries to infer what inheritance patterns and modifiers are responsible for the production of the observed phenotypes in a population of organisms generated by the computer. The SOLVER program does not have access to the genetics parameters that were defined in the customization section of the GENERATOR. The SOLVER is therefore an expert at making inferences from phenotypic data and recommending crosses when requested.

Major Components of the MENDEL System

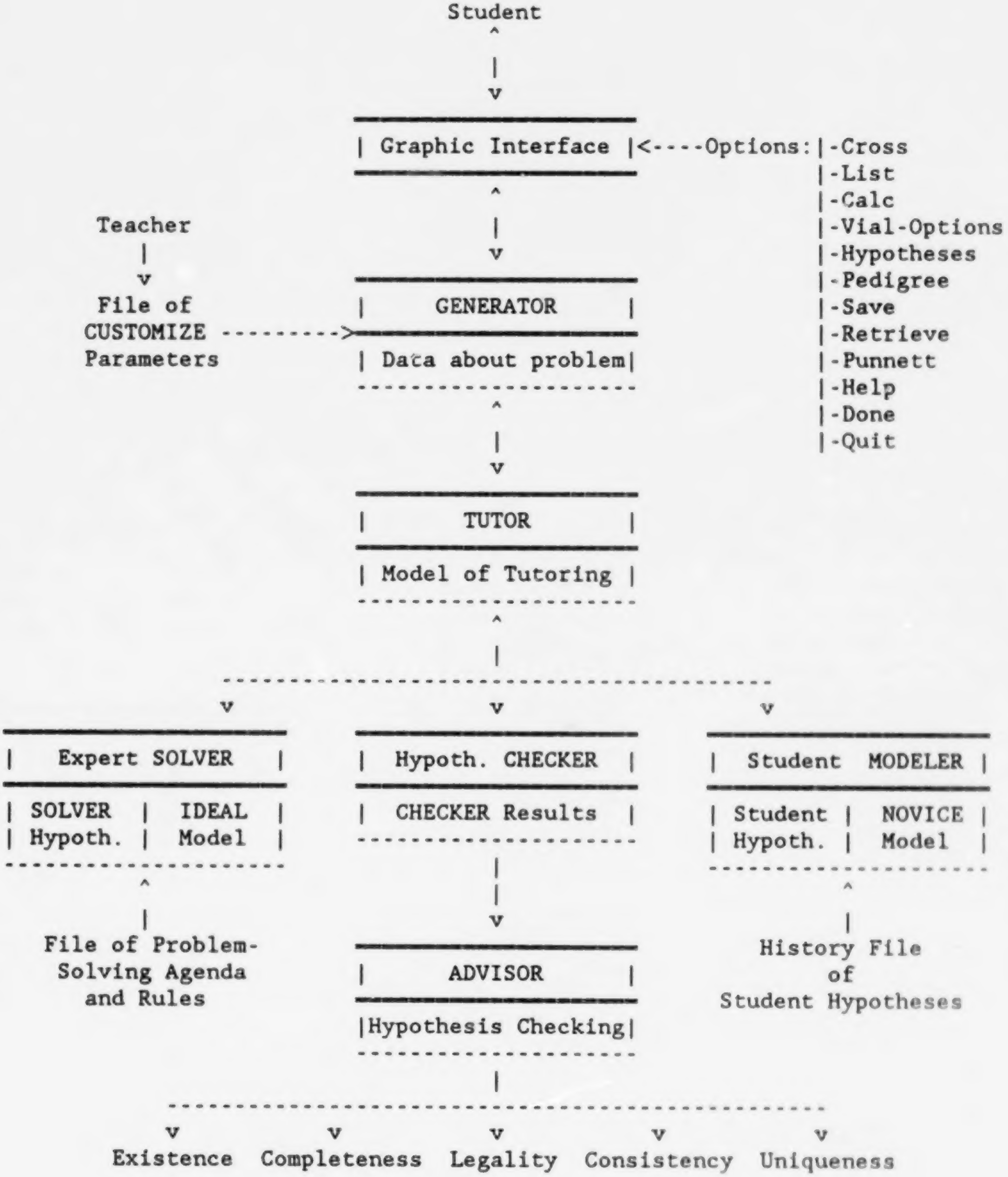

Student
Graphic Interface
Options:
-Cross
-List
-Calc
-Vial-Options
-Hypotheses
-Pedigree
-Save
-Retrieve
-Punnett
-Help
-Done
-Quit
Teacher
File of CUSTOMIZE Parameters
GENERATOR
Data about problem
TUTOR
Model of Tutoring
Expert SOLVER
SOLVER Hypoth.
IDEAL Model
Hypoth. CHECKER
CHECKER Results
Student MODELER
Student Hypoth.
NOVICE Model
File of Problem-Solving Agenda and Rules
ADVISOR
Hypothesis Checking
History File of Student Hypotheses
Existence
Completeness
Legality
Consistency
Uniqueness

Figure 2

Sample GENERATOR Screen

```
GENERATOR: Cross List Vial-opts Pedigree Hypoth Save Retrieve Done Exit
Cross any male and female organism from any vial.

-----vial-A------     -------------------vial-A----------------------
| 1mBH    4fBH  |     |Parents: Initial Population                   |
| 1mBR    1fBR  |     |Traits:  EYES EARS                            |
| 2mBN    1fBN  |     | Sex   #     EYES     EARS                    |
| 4mYH    3fYH  |     ------------------------------------------------
|               |     |  m    1     BLUE     HUGE                    |
----field-pop----     |  f    4     BLUE     HUGE                    |
                      |  m    1     BLUE     ROUND                   |
                      |  f    1     BLUE     ROUND                   |
                      |  m    2     BLUE     NORMAL                  |
                      |  f    1     BLUE     NORMAL                  |
                      |  m    4     YELLOW   HUGE                    |
                      |  f    3     YELLOW   HUGE                    |
                      ------------------------------------------------
```

Figure 3

The SOLVER is guided by a problem-solving Agenda. This **Agenda** was derived from our research on geneticists (Collins, 1986) and is summarized in an abstract form in Figure 4. The **Agenda** has been structured as a problem-solving tree in order to facilitate computer processing and consists of a series of **methods** (leaf nodes) and **strategies** (non-leaf nodes). A **method** is a LISP function which is called explicitly by name and immediately performs a required action. A **strategy** consists of an ordered list of steps (either strategies or methods) followed by a test for an end condition which may or may not be satisfied. For example, **Experiment** in Figure 4 is a strategy that is cycled through by the SOLVER until it is satisfied that an hypothesis is consistent with the data to an arbitrary degree of confidence.

The goal of the SOLVER is to find inheritance patterns and modifiers and thus account for the entire phenotypic data set that has been generated by the student. Hence, each of the following steps in the **Agenda** work towards that end (items with a * besides them are methods):

*1. **Redescribe Initial Data** is the process of transcribing the "physical" data a student might see (such as trait names, variations associated with traits, and numbers of individuals) into a LISP Frame that the SOLVER can manipulate.

*2. **Plan Initial Crosses** examines the initial redescribed data and plans crosses to new data.

3. **Experiment** is the core process of the SOLVER. The SOLVER repeatedly makes and tests hypotheses until it becomes satisfied with an hypothesis. The hypotheses manipulated within **Experiment** are 'within-trait' hypotheses (i.e., independent of hypotheses of other traits). Across-trait hypotheses such as pleiotropy and linkage are dealt with in **Check Results**.
    *a. **Generate Hypotheses** defines a specific class of hypotheses that the SOLVER checks against the phenotypic data. A class is formed by choosing an inheritance pattern (IP) and deciding whether sex linkage or lethality occur in the population.
    b. **Test and Refine Hypothesis** is the process of interpreting data from a cross and determining how well it fits each trait's current class of specific hypotheses.
        *i. **Pick Cross** examines the crosses that have been planed and picks one to be done.
        *ii. **Redescribe Cross Data** is the process of transcribing data from the last cross into the SOLVER's constructs.
        *iii. **Plan Future Crosses** chooses which individuals to cross to obtain new data.
        *iv. **Explain Cross** is a pattern matching process that uses chi-square analyses. For each trait's class of specific hypotheses, an algorithm is applied to determine if a cross fits a specific hypothesis, and, if so, how well.

MENDEL's Problem-Solving Agenda

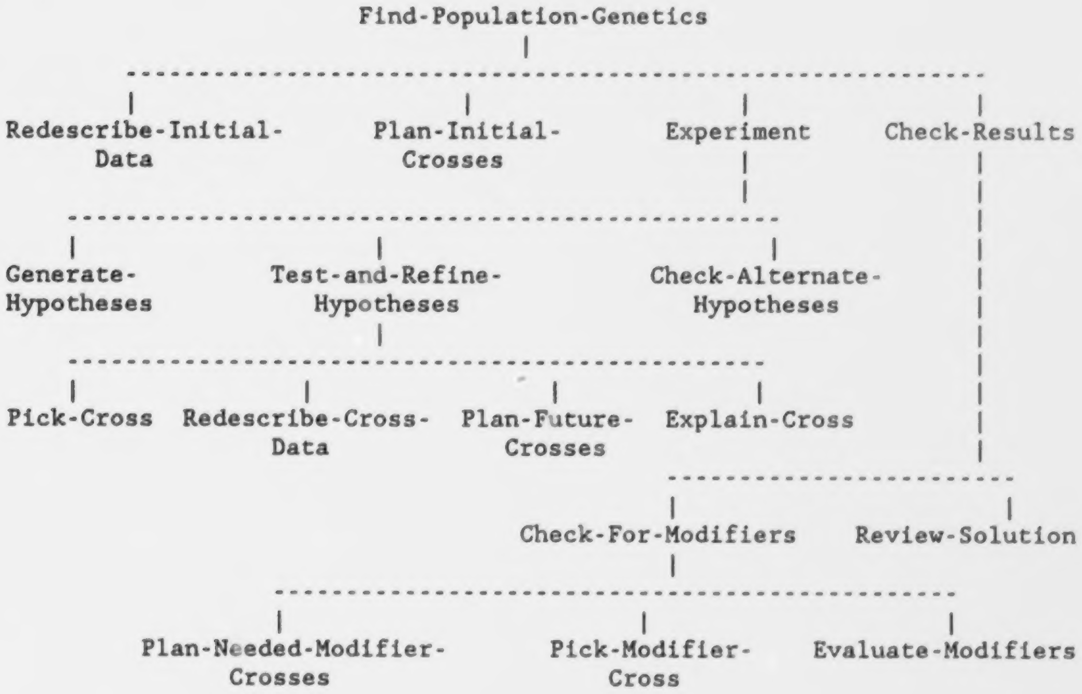

Find-Population-Genetics
Redescribe-Initial-Data
Plan-Initial-Crosses
Experiment
Check-Results
Generate-Hypotheses
Test-and-Refine-Hypotheses
Check-Alternate-Hypotheses
Pick-Cross
Redescribe-Cross-Data
Plan-Future-Crosses
Explain-Cross
Check-For-Modifiers
Review-Solution
Plan-Needed-Modifier-Crosses
Pick-Modifier-Cross
Evaluate-Modifiers

Figure 4

- *c. **Check Alternate Hypotheses** examines the possibility that there exists an hypothesis other than the one generated by **Generate Hypotheses** that can also explain the data. Relevant information discovered here is made available to **Generate Hypotheses**.

*4. **Check Results** checks for across-trait modifiers and then reviews the solution.
- a. **Check For Modifiers** checks for across-trait modifiers (i.e. linkage and pleiotropy).
  - *i. **Plan Needed Modifier Crosses** plans the crosses that are needed in order to test for across-trait modifiers.
  - *ii. **Pick Modifier Cross** picks a cross that was planed to test for across-trait modifiers.
  - *iii. **Evaluate Modifiers** examines the data to see if a test for across-trait modifiers can be made. If the needed data exists, then the test is made.
- *b. **Review Solution** checks for correctness of the solution.

In summary, the SOLVER uses a rule-based approach to generating hypotheses about inheritance patterns and to recommending crosses within the constraints of these hypotheses. It utilizes data from student crosses although it is capable of recommending further crosses. It has the ability to keep track of its own inferences and the ability to find the best data-fitting hypothesis about inheritance patterns and modifiers. The SOLVER is explained in greater detail in Appendix G.

The Interface Options of the GENERATOR Component of MENDEL. Our research on how students use the GENERATOR program revealed that many of them did not think in generational terms when they solved genetics problems (Slack, 1988). That is, they used organisms from the initial population as parents for all or most of their crosses. This indicated to us that a notational system which summarized existing data across generations would help students think in trans-generational terms. We therefore designed the **Pedigree Diagram** option and added it the MENDEL system.

While designing the **Pedigree** option, we also added a facility to the interface part of the GENERATOR component in order to let students enter possible genotypes for each parent organism and class of offspring. This was intended to reinforce the idea that more than a single genotype to phenotype mapping was possible at any point in the solution. By entering possible genotypes on the **Pedigree Diagram**, students were thus encouraged to use information from a cross as a basis for selecting the parents of future crosses. In this way, they would begin to see crossing as a procedure for producing knowledge about the population.

Our work on student use of the MENDEL system then led us to propose several fundamental changes in our approach. First, we needed to develop notational systems in addition to **Pedigree** in order to help students grasp other key aspects of solving problems. Second, we needed a way to have students enter their hypotheses into the MENDEL system. And finally, we needed to connect a student's understanding of the problem-solving task with their use of these notational systems. We therefore created:

1. **Hypothesis-Entry Facility**: to collect data about student hypotheses (See

Figure 5 below):

   a. **Hypothesis Data-Entry Mode**: to permit the student to enter their hypothesis;
   b. **Current-Hypothesis Summary Form**: to serve as a data-management summary of the student's hypotheses;
   c. **Chromosome-Pair Graphs**: to summarize a student's hypotheses about inheritance patterns and modifiers at the chromosomal level;
   d. **Expression Charts** (not shown in Figure 5): to display the complete genotype-to-phenotype relationships for any inheritance pattern and modifiers for any trait;

2. **Punnett Squares**: to display certain subsets of genotype-to-phenotype relationships according to the principles of segregation and independent assortment as well as to collect data on what the student believed was interrelated. Hence, there are two uses of the **Punnett Squares**:
   a. genotype-to-phenotype mapping with respect to actual data;
   b. "scratchpad mode" that is not tied to actual data;

3. **Cross Equations**: to represent Cross possibilities for each inheritance pattern.

4. **Inquiry Strategy Tree**: to display the overall inquiry strategy that the expert SOLVER uses to solve the same genetics problem as the student (See Figure 4). This includes problem-solving rules and justification heuristics.

These features therefore reflected a basic reorientation. We would hence-forth be tutoring students to structure their knowledge about the genetics problem-solving domain and to build up a scientific strategy similar to the problem-solving agenda in our program. Tutoring would be more like coaching and scientific inquiry would be focused on model-based problem solving in transmission genetics.

The **Hypothesis-Entry** facility of the GENERATOR component and associated notational systems helped us address a number of issues that we had found to be important in our research -- namely, that students: did not generate hypotheses; did not relate genotypes to phenotypes; and, did not relate their problem solving to the events of meiosis (Stewart, 1983; Albright, 1987; Slack, 1988). It also became clear from our research with high-school and college students that many of them did not consider more than one explanation for their data. This lack of perspective occurred at both the hypothesis-generation level and at the cross-plan level. Students therefore needed some way to recognize the possibility that there might be multiple explanations for cross results without continually being told about this possibility. The notational systems, therefore, structured the representation of genetics objects, processes, and states so that these possibilities would be obvious.

Sample Hypothesis Entry GENERATOR Screen

```
GENERATOR(Hypoth): SD CD MA IntrAct | SeX-Lnk Lethal PleiO LinKd PeNetr IntrFer
    Remove-hypoth  Graph-change  Express-chart  Link-Dist  Pedigree  Help  Quit
Add SeX-Linkage modifier for a trait.

....Data Entry..........Current Hypothesis............Chromosome Pair Graph....
Choose Trait:     . EYES( ) IP: CD                .       || ||          || ||
 EYES       ( )   .   Mod(s):                     .       || ||        X || || Y
 ANTENNAE   ( )   .                               .       || ||          || ||
 OCELLI     ( )   . ANTENNAE( ) IP: SD            .       || ||          ||
 BODY       ( )   .   Mod(s): sex-linked          .     1 || ||        2 ||
                  .                               .       || ||          ||
                  . OCELLI( ) IP: MA (3 alleles)  .       || ||          ||
Type 1st letter   .   Mod(s):                     .       || ||          ||
 or <END> key     .                               .       || ||          ||
   to quit        . BODY( ) IP: Interaction       .     Pair  A        Pair  B
                  .   Mod(s): <<linkage( - )>>    .       || ||          || ||
Is entry OK? y    .                               .       || ||          || ||
                  .                               .       || ||        4 || ||
                  .                               .       || ||          || ||
                  .                               .     3 || ||          || ||
                  .                               .       || ||          || ||
                  .                               .       || ||        5 || ||
                  .                               .       || ||          || ||
                  .                               .       || ||          || ||
                  .                               .     Pair  C        Pair  D
```

Figure 5

For example, the **Chromosome Graph** shows the inter-relationship between a content-domain model and a potential problem-solving activity (See the bottom right portion of Figure 5). If a student enters an hypothesis that the first trait is due to simple-dominance with sex-linkage, he or she immediately sees a graphic representation of this hypothesis on the **Chromosome Graph**. If the student then proceeds to hypothesize that the second trait is linked to the first (by entering this hypothesis via the **Hypothesis-Entry** facility), he or she will immediately see a consequence on the **Chromosome Graph** (ie. the second trait also becomes sex linked). He or she can then test for this consequence by performing appropriate crosses. The **Chromosome Graph** allows students to observe a pictorial representation of how the current hypotheses are related to the arrangement of loci on chromosomes and to consider, where appropriate, the relationship of meiosis to problem solving. This kind of representation makes basic knowledge about the domain accessible to students as they solve genetics problems and also lets them immediately see the consequences of their hypotheses at the appropriate knowledge levels.

The TUTOR Component of MENDEL. The TUTOR is the central organizing mechanism in MENDEL. The TUTOR, at present, is simply a control loop that invokes the Hypothesis-CHECKER when the student presses the **Done** key. However, the TUTOR is capable of incorporating other specific tutoring mechanisms because it is based on the concept of Post-Socratic tutoring first formulated by Jungck and Calley (Jungck & Calley, 1985). Before describing our hypothesis-checking mechanism, we will briefly describe the philosophy of Post-Socratic tutoring.

Post-Socratic tutoring provides a framework for the interactions between the computer-tutor and the student. This framework claims that students learn best by doing and by immediately seeing the consequences of their actions. We have extended this philosophy by providing a number of notational systems that display the consequences of student actions and hypotheses on various levels of genetics knowledge (phenotypic, genotypic, and meiotic). We deliberately chose to develop notational tools before we addressed specific tutorial interventions. Hence, in the MENDEL system, we represent genetics problems in various ways so that the consequences of student hypotheses and actions are made explicit on various levels of interconnected knowledge. We believe that the constant intervention of a tutor is not needed in this situation. Rather, what is needed is a TUTOR that solves the problem along with students and helps them to structure the problem so that they can eventually solve it on their own.

The Post-Socratic philosophy outlined above has helped us create a problem-solving environment that encourages students to work through their own difficulties and only receive tutorial advice when they are at an impasse. This philosophy has also helped us recognize a new student-teacher relationship. In traditional classrooms, problem solving is structured so that the tutor knows the final answer as well as the solution paths to that final answer. This turns problem solving into a rational reconstruction of scientific problem solving. In MENDEL, on the other hand, the tutor and student are equals in terms of their knowledge of the solution to the specific problem. Students must therefore take an active part in the important decisions required for scientific problem solving. Tutors are helpful on the basis of their problem solving, genetics, and advising expertise and not on the basis of their prior knowledge of the answer.

Post-Socratic tutoring fits well into the mainstream of Intelligent Tutoring Systems (ITS) research that is currently under way. For example, Collins and his colleagues have argued that the tutoring strategies of human tutors can be modeled if these strategies are related to the discrepancies between a student's knowledge of a domain and the facts and underlying mechanisms of that content domain (Stevens, Collins, & Goldin, 1982; Collins & Stevens, 1983). In the MENDEL system, we pursue a similar strategy that addresses discrepancies between the student's hypothesis about the underlying inheritance patterns and modifiers and the actual phenotypic data. Rather than intervening with textual discourse to address such discrepancies, however, we offer the student a number of notational systems that structure their problem space in such a way as to make these discrepancies more obvious. This is intended to accommodate as wide a range of student problem-solving styles as possible.

Clancey's work on extending the MYCIN system provides another example of an important component of successful tutoring. MYCIN was developed to help medical personnel diagnose bacterial infections (Buchanan & Shortliffe, 1984). Clancey and his colleagues then added a tutorial component to the MYCIN system and found that users needed to "look" into the logic of the system (Clancey, 1986a, 1986b). He therefore developed a number of graphic facilities called GUIDON-WATCH that provided users with multiple windows into the reasoning of his system (Richer & Clancey, 1985). Someone who is learning to solve problems in his domain, he argued, needed an explicit representation of the evidential network during medical diagnosis and a representation of the processes of problem solving. We have taken this idea in MENDEL and formulated a number of traditional as well as new notational systems to help students get an explicit view of their own knowledge, their hypotheses, and the genetics knowledge involved in the problem (e.g., traditional: **Punnett Square**, **Pedigree Diagram**; new: **Expression Chart**, **Chromosome Diagram**, and **Inquiry Strategy Tree**). We have, in fact, developed an **Hypothesis-Entry** facility that helps students reformulate their tentative hypotheses into a computable form by employing these notational systems.

Finally, several researchers have argued that tutoring systems must address qualitative reasoning processes for there to be successful model-based learning (Stevens & Collins, 1980; Gentner & Stevens, 1983; White & Fredrickson, 1986). Hence, they show that experts are successful at problem solving because they have a mental representation of some content domain which they can picture and "run" in their imagination. Novices have trouble solving problems in these content domains because they lack, or have incorrect, mental models. In MENDEL, we conjecture that students will develop a robust understanding of genetics problem solving if they are given model-oriented notational systems that represent genetics objects, processes, and states, and if they are given model-oriented advice as they solve problems. An example of the latter would be a series of general to specific messages that direct the student's attention to the discrepancies between their hypotheses about the current problem and the actual data that they see. Hence, we conjecture that if we make an expert's representation of the knowledge and problem-solving procedures cognitively accessible to students by the manner in which we structure the representation of the problem and by the way we have them enter their hypotheses, we will encourage genetics understanding in students. In the future, we will add one final component to the MENDEL system: justification of the solution in model-based terms. Our problem-solving environment, our notational tools, and our problem-solving advice will all provide students with a number of representations of knowledge (content and inquiry knowledge) that they do not initially have but that will encourage them to develop their own models of genetics problem solving.

The Hypothesis CHECKER. We begin by discussing the hypothesis-checking strategy. Although the hypothesis-checking mechanism has been integrated into the MENDEL system, many of the research questions that it raises are still under investigation. We will therefore describe the hypothesis CHECKER, while acknowledging that it will undoubtedly undergo revision as the appropriate research is concluded.

We have found that genetics tutors help students check their hypotheses against the data by asking probing questions about the consistency, completeness, and goodness-of-fit of their hypotheses (Stewart, 1987b). This general tutoring strategy served as one well-defined strategy in our system and helped us determine what student actions were significant indicators of the presence or absence of hypothesis-checking. This tutoring strategy also formed the guidelines for our Hypothesis-CHECKER. The TUTOR, in effect, acts as a control driver that decides what to do and when to do it. The CHECKER, on the other hand, evaluates the student's hypothesis against the data for existence, completeness, legality, consistency with the data, and goodness-of-fit. The CHECKER then returns these results back to the TUTOR which in turn drives the ADVISOR to deliver advice to the student about their hypothesis. These conditions of the hypothesis CHECKER are described below:

1. the **existence** of a student's hypothesis that links genotype to phenotype information for each trait in a problem;
2. the **completeness** of a student's hypothesis that bridges genotype to phenotype information;
3. the **legality** of a student's hypothesis with respect to the definitions of genetics concepts about objects, processes, and states;
4. the **consistency** of a student's hypothesis with respect to the phenotypic data;
5. the **goodness-of-fit** of a student's hypothesis (i.e., which legal and complete hypotheses fit the data);
6. the **checking** of whether common modifiers have been accounted for.

The Student-MODELER Component of MENDEL. The function of the Student-MODELER is to provide the TUTOR with information about the problem-solving performance of students and the conceptual basis of this performance. In order to do this, the MODELER must contain a data-base of specific student actions and interpretations of the current data as well as a knowledge-base of typical student conceptions and misconceptions. The former type of information is all that is currently needed for the operation of the hypothesis-checking mechanism. The latter type of information will come from our current research (Slack, 1988; Albright, 1987; Stewart & Dale, 1987; Stewart, 1987) in order to support other types of tutoring mechanisms.

Our current research on student actions and misconceptions is guided by the major problem-solving steps of the entire range of the inquiry process (See Figure 4). Hence, we are asking questions within the framework of MENDEL's problem-solving Agenda.

The Notational Systems of the GENERATOR Component of MENDEL. The Hypothesis-Entry facility which we described earlier helped us gather data about student thinking as they solve GENERATOR-created problems. However, the associated notational systems also served several other functions besides data collection such as: a communication link between the student and the TUTOR; multiple windows into the knowledge-structures and logic of MENDEL; and, multiple ways of showing students the consequences of their hypotheses (so that they could pose "what if" questions as they solve problems).

We are currently investigating how the notational systems might serve as cognitive tools to help students think about genetics problem solving in model-based terms. We are studying the role of the notational systems both individually and collectively with high-school and university students as they solve genetics problems. We are also investigating the extent to which such students use these notational systems and the effect these notational systems have on their understanding of genetics problem solving. Specifically, we are asking: about the usefulness of each notational tool; about which tools are used and when in the problem-solving process; and, about how students use the notational systems to justify their interpretation of the data. We are also gathering information on the reaction of instructors to the Hypothesis-Entry interface.

The Problem-Solving ADVISOR Component of MENDEL. Aside from providing advice that results when the CHECKER examines the student's hypothesis, the ADVISOR will also help a student justify their solution to a problem. Justification is difficult to operationalize and transfer into a computer environment (See Appendix A). We therefore anticipate that the future MENDEL system will only be able to provide examples of justified solutions (using the notational tools that we have developed) and not be able to evaluate the justifications of students. We will nevertheless direct our research efforts into how experts and novices justify their solutions to genetics problems and how human tutors help novices justify their solutions in the hope that some aspects of the justification process can be routinized.

Since our problem-solving agenda and our notational tools have provided such a useful framework for all of our previous research and development, we are using these ideas as a framework for our research into justification heuristics. That is, we are using the sub-stages of our problem-solving agenda (See Figure 4) to categorize the problem-solving process and investigate which genetics justifications are germane to each stage of the process. We are also using our knowledge of the notational systems to probe a student's understanding of the connection between genetics objects, processes, and states with the problem-solving process at that stage. Since the notational tools represent genetics knowledge at the symbolic level, we are looking to see if the students' justifications are clear at the abstract symbolic level as well as whether they can explain genetics phenomena in terms of meiotic models. Appendix F contains an example of an ideal justification as it emerges under the probing questions of a human tutor.

Finally, we are investigating what specific advice is appropriate for the tutoring strategies as they emerge. For example, we would like to know how best to review and analyze the history of a student's problem-solving session. We already have a mechanism in MENDEL for determining what actions an expert solver would take in response to the student's data at each stage of the problem-solving process.

E. Project Results

As described above the MENDEL system consists of a problem GENERATOR, an expert SOLVER, interface options, an hypothesis CHECKER, a student MODELER, and an ADVISOR. It was also indicated in the main body of the report that we have not, to date, accomplished as much as we set out to accomplish, having instantiated an hypothesis CHECKER, but not other tutorial strategies. In the Project Results section we will describe: the effects of our project on the participants; the dissemination efforts, and; the evaluation of the project.

The Effects of the Project on the Participants. A significant effect of the project, which will continue beyond the FIPSE supported phase of the IGT project, is the establishment of a cooperative atmosphere among our department (Curriculum & Instruction) and the departments of Computer Science, Zoology and Genetics. This is significant in as much as these departments and ours are in three different Colleges (Education, Letters and Science and Agriculture) that have not, in the past, interacted to any great extent. The interactions with the computer science department have primarily been with Professor Larry Travis. This has occurred because two of his graduate students were employed on the IGT project. The interactions with Zoology and Genetics have been with faculty members who have acted as genetics experts in our efforts to build the SOLVER and because they have used the IGT software (the GENERATOR) with courses or workshops that they have conducted. Professor Raymond Kessel of the Genetics Department has been most enthusiastic in using our software.

In addition the IGT project has had an effect on a national biology education project as we have become involved in BioQUEST, a consortium of six universities, that is developing a university freshmen biology course around 12 biology simulation programs. Although the entire MENDEL system will not be a part of BioQUEST, the results of our project have been used to shape the development of a MacIntosh version of the GENERATOR, and to add to the overall philosophy of BioQUEST. This link will allow us to influence a national software development project that has the potential to have widespread influence on biology education.

The above have been the most direct effects of the project in terms of the initial target audience. However, there have been effects on other audiences. Most significant has been the number of high school biology teachers that we have worked with in summer institutes that are jointly sponsored by the Departments of Curriculum & Instruction and Genetics. During the last two years we have worked with over 100 such teachers. While we have shown them the entire MENDEL system, it is the GENERATOR that is of most interest to them. By using the generator and the issues that we have developed around the SOLVER, hypothesis CHECKER, interfaces, MODELER, and ADVISOR we have been able to: help them to think about their genetics knowledge and problem-solving strategies in new ways; and, help them to think about students genetics learning and problem solving genetics in new ways. Because of this successful application of the results of our project we will continue working with high school teachers and prospective biology teachers as they obtain teaching certifications. For example, we are in the process of designing a course for biology certification students that will make extensive use of MENDEL and the instructional ideas that are embedded within it.

Dissemination of Results. The dissemination of the results have to this point, in addition to the involvement with high school teachers and the BioQUEST project, included paper presentations, journal articles and technical reports. We have made over twenty presentations related to IGT at local, state, and national meetings, including those of the American Educational Research Association, the National Association for Research in Science Teaching, the Third Conference on Artificial Intelligence and Education, and the Association for Educational Communications and Technology. Publications include one in the Journal of Machine-Mediated Learning, and a second that is in press at Science Education. In addition, we have produced five technical reports form the IGT project and have several more in progress.

We are now in the process of making final arrangements for the dissemination of one aspect of the IGT project, the problem GENERATOR. We are removing the GENERATOR for the larger MENDEL project and will disseminate it nationally through the University of Wisconsin WISCware organization. By doing this, high school and university biology teachers will have an simulation program that will run on a minimally configured IBM. By making the GENERATOR available in this format the philosophy that underlies the IGT project will be disseminated more widely than can the MENDEL program.

Evaluation. The evaluation of IGT has included evaluation of the component parts, particularly the GENERATOR, SOLVER, and hypothesis entry facility for their fidelity to genetics. This has been done by involving geneticists as evaluators. Professor John Jungck, the biology department chair at Beloit College and a national leader in genetics education and biocomputing, has been a paid consultant on the project. In addition members of the Genetics and Zoology departments on the Madison campus have acted as consultants.

In addition, the GENERATOR has been extensively used with both university and high school students. This work has served as a formative evaluation that has provided us with insights for revising the GENERATOR to make it more convenient for students to use. The results of these evaluations can be found in Slack and Stewart (1988a; 1988b).

Because of the unanticipated length of time that it took to develop the GENERATOR, the expert SOLVER, and the hypothesis entry facility we have been unable to do a summative evaluation of the entire MENDEL system. This has been one of the major disappointments of the project. We have submitted a proposal to NSF to allow us to complete work on MENDEL including a thorough evaluation of student use of it. We expect to begin this continued phase of the project in the Fall of 1988.

F. Project Summary and Conclusions

In this section we will discuss insights gained from work on the IGT, including those in which our original ideas, plans and goals have changed. By doing this we hope that others may benefit.

Some of the most important insights have been personnel ones about genetics and the teaching of genetics which now influence our thinking outside the IGT project. Foremost most among these have been insights about the organization and structure of classical genetics that is necessary to facilitate model-based problem solving. We have found it interesting that these insights are only rarely made explicit in university genetics instruction. We have been heartened by the positive acceptance of these ideas, particularly about what constitutes core information in genetics. A second insight of a related nature has been a deepened sense of conviction of the importance of having students (and instructors) solve realistic genetics problems of the type produced by the GENERATOR. By solving such problems there is an enhanced opportunity to engage in what the 3P's of science education (Peterson & Jungck, et.al., 1988) -- problem posing, problem solving, and persuasion of peers. By engaging in the 3 P's it is our conviction that students will learn conceptual knowledge of genetics, problem-solving strategies independent of and specific to genetics, and insights into the nature of science as a human problem-solving, decision-making activity.

A second category of insights relate to our growing understanding of the role that tutoring systems, specifically MENDEL, might have for classrooms in the immediate future. Early in the project we realized the difficulty of specifying all of the possible conditions and actions that a tutor might take in our unstructured GENERATOR's problem-solving environment. Therefore we have focused more on helping students structure their genetics knowledge and on coaching students on their problem solving and experimentation activities. Hence, we have only instantiated an hypothesis-checking tutorial strategy in our computer environment. However we have developed tools that help students structure their genetics knowledge, constructed from repeated problem solving experiences. We feel that this is more fundamental than helping them learn from an expository tutoring approach. What all of this has meant is that we have come to be much more concerned about the role of course instructors in the MENDEL system. From this insight we have now begun to consider what types of advice a computer can best give (it is likely to be on actions) and what types a human tutor is going to be best able to provide (most likely advice in the form of justifications for actions).

Another insight that we have gained that should be of value to others who might be interested in our software or in developing software with related goals is the need to spend extended periods of time with potential users of the software. This is for several reasons, not the least of which this type of software is much more sophisticated than typical educational software and for potential users to become convinced of its utility requires, we think, actual experience with the software and software experts. Insights about the software are less likely to emerge from using it with a users manual than from workshop experience with someone very familiar with the software. The second reason for the value of this extended experience is the importance of the underlying philosophy of post-Socratic tutoring, learning by doing, etc to the use of the software by students. We feel that this will require that potential users have the opportunity to interact with members of the IGT project concerning the philosophy. It is in these workshops that serious consideration can be given to how the MENDEL software is to be integrated into a course. It is not likely to be a success if it is simply treated as an "add on".

REFERENCES

Albright, W. (1987). Problem-solving Performance of University Biology Majors on Realistic Genetics Problems. Master's Thesis. The University of Wisconsin-Madison.

Buchanan, B.G. & Shortliffe, E.H. (Eds.). (1984). Rule-Based Expert Systems: The MYCIN Experiments. New York: Addison-Wesley.

Clancey, W.J. (1984). Methodology for building an intelligent tutoring system. In W. Kintsch, J.R. Miller, & P.G. Polson (Eds.), Method and Tactics in Cognitive Science. Hillsdale, N.J.: Lawrence Erlbaum & Associates. pp. 51-83.

Clancey, W.J. (1986a). From Guidon to Neomycin to Heracles in twenty short lessons. AI Magazine, August, pp. 40-60, 187.

Clancey, W.J. (1986b). Heuristic classification. In J.S. Kowalik (Ed.), Knowledge-Based Problem-Solving. Englewood Cliffs, NJ: Prentice-Hall.

Clancey, W.J. (1987). Knowledge-Based Tutoring: The GUIDON Program. Cambridge, MA: MIT Press.

Collins, A. (1986). Strategic Knowledge Required for Desired Performance in Solving Transmission Genetics Problems. Ph.D. Dissertation. The University of Wisconsin - Madison.

Collins, A., & Stevens, A.L. (1983). A cognitive theory of inquiry teaching. In C.M. Reigeluth (Ed.), Instructional-Design Theories and Models. Hillsdale, NJ: Lawrence Erlbaum & Associates.

Frederickson, N. (1984). Implications of cognitive theory for instruction in problem solving. Review of Educational Research, 54(3), 363-407.

Gentner, D. & Stevens, A. (Eds.). (1983). Mental Models. Hillsdale, N.J.: Lawrence Erlbaum Associates.

Jungck, J.R., & Calley, J.N. (1985). Strategic simulations and post-socratic pedagogy: Constructing computer software to develop long-term inference through experimental inquiry. The American Biology Teacher, 47(1), 11-15.

Jungck, J.R., & Calley, J.N. (1986). GENETICS: Strategic Simulations in Mendelian Genetics. Wentworth, NH: COMPRESS.

Kulik, J.A., & Bangert-Drowns, R.L. (1983). Effectiveness of technology in pre-college math and science teaching. Journal of Educational Technology Systems, 12(2), 137-158.

Kulik, J.A., Kulik, C.C., & Cohen, P.A. (1980). Effectiveness of computer-based college teaching: A meta-analysis of findings. Review of Educational Research, 50(4), 525-544.

Larkin, J. McDermott, J., Simon, D.P., & Simon, H.A. (1980). Expert and novice performance in solving physics problems. Science, 208, 1335-1342.

National Academy of Science and National Academy of Engineering (1982). Science and Mathematics in the Schools: A Report of a Convocation. Washington, DC: National Academy Press.

Slack, S. (1988). Problem-solving strategies used by high school students to solve realistic problems. Master's Thesis. University of Wisconsin - Madison.

Peterson, N., & Jungck, J. (1988). Problem-posing, problem-solving and persuasion in biology education. Academic Computing, March/April, 19-20.

Richer, M.H., & Clancey, W.J. (1985). Guidon-watch: A graphic interface for viewing a knowledge-based system. IEEE CG&A, November.

Stevens, A.L., & Collins, A. (1980). Multiple conceptual models of a complex system. In R.E. Snow, P. Federico, & W.E. Montague (Eds.), Aptitude, Learning and Instruction. Hillsdale, NJ: Lawrence Erlbaum Associates.

Stevens, A.L., Collins, A., & Goldin, S.E. (1982). Misconceptions in students' understanding. In D. Sleeman & J.S. Brown (Eds.), Intelligent Tutoring Systems. London: Academic Press, pp. 13-24.

Stewart, J. (1983). Student problem solving in high school genetics. Science Education, 67(4), 523-546.

Stewart, J. (1988). Potential learning outcomes from genetics problem-solving: The type of problem makes a difference. Science Education. In press.

Stewart, J., & Dale, M. (1987). High school students' understanding of chromosome/gene behavior during meiosis. In review, Science Education.

Stewart, J., & Slack, S. (1987). A description of student problem-solving on realistic genetics problems. In Proceedings of the Second International Seminar: Misconceptions and Educational Strategies in Science and Mathematics.

Stewart, J., & Slack, S. (1988). High-school students problem solving on realistic genetics problems. In review Journal of Research in Science Teaching.

Streibel, M.J., Stewart, J.H., Koedinger, K., Collins, A., & Jungck, J. (1987). MENDEL: An intelligent computer tutoring system for genetics problem solving, conjecturing, and understanding. Machine-Mediated Learning, 2(1&2), 129-159.

White, B., & Fredriksen, J. (1986). Progressions of Qualitative Models as a Foundation for Intelligent Learning Environments. Technical Report No. 6277. Cambridge, MA: Bolt, Beranek, and Neuman.

APPENDIX A

MENDEL Paper

MENDEL: AN INTELLIGENT COMPUTER TUTORING SYSTEM FOR GENETICS PROBLEM-SOLVING, CONJECTURING, AND UNDERSTANDING*

Michael J. Streibel

Educational Technology Program
University of Wisconsin
Madison, Wisconsin

James Stewart

Curriculum and Instruction Department
University of Wisconsin
Madison, Wisconsin

Kenneth Koedinger

Carnegie-Mellon University
Pittsburgh, Pennsylvania

Angelo Collins

Curriculum and Instruction Department
Kansas State University
Manhattan, Kansas

John R. Jungck

Biology Department
Beloit College
Beloit, Wisconsin

---

*An earlier version of this paper was presented at the 1986 annual meeting of the *American Educational Research Association*.

Machine-Mediated Learning, Volume 2, No. 1 & 2, 129–159, 1987
0732-6718/84/01029-00$3.00/0
Copyright © 1987 Taylor & Francis, and Mentor Systems, Inc.

Abstract This paper describes an advice-giving computer system for genetics education called the **MENDEL** system that is based on research in learning and genetics problem solving as well as on recent advances in expert systems. The **MENDEL** system is designed to help students gain a better understanding of genetics and scientific inquiry by providing them with the opportunity to solve realistic genetics problems and obtain tutorial assistance that is tailored to their genetics knowledge and level of proficiency at problem-solving. **MENDEL** consists of a problem GENERATOR component and a TUTOR component. The TUTOR includes: a rule-based, expert SOLVER; a problem-solving ADVISOR; a student MODELER; and, a video/graphics LIBRARIAN.

Introduction

There is a growing literature in education and psychology that addresses the need for open-ended problem-solving in science education [1, 2]. There is also an increasing call for the instructional use of microcomputers in science education (as seen in the pages of *The American Biology Teacher* and*The Science Teacher*). Finally, there is an emerging discipline within artificial intelligence research that deals with the design and use of intelligent tutoring systems and advice-giving systems [3, 4]. These trends are converging so that the time is right to bring the theoretical and practical advances within each discipline to bear on the design and use of computers in science education. For example, research in education and psychology has focused on: student alternate conceptions [5-8]; problem-solving [9, 10]; and teaching for conceptual change [11,12]. Research in artificial intelligence, on the other hand, has focused on: the development of knowledge representation schemes (e.g., frames, production rules, semantic networks, etc.), the design of intelligent tutoring systems [13, 3, 14-17], and the instructional potential of intelligent tutoring systems [18-20]. These developments complement and reinforce each other so that educational software can now be based on theories of teaching, learning and problem-solving[13].

For the past several years, we have carried on a research and development effort that has focused on promoting improvements in teaching genetics at the high-school and college levels. This work has entailed the analysis of high school students' knowledge of transmission genetics as well as how their knowledge influences their problem-solving performance [21-23]. More recently, we have been studying the strategies that beginning university students [24], high school students [25] and geneticists [26] use to solve realistic genetics problems generated by a microcomputer.

We have also developed genetics simulation programs [27] that allow students to act like genetics researchers. These programs, called strategic simulations, provide students with the opportunity to develop problem-solving skills and long-range research strategies similar to those used by transmission geneticists [28,29]. Finally, we have been involved with the rapidly developing technology of interactive videodiscs [30] and the critical analyses of the use of computers in education [31].

Drawing on our own interests and research as well as on the recent research on expert systems [3, 32, 33, 17, 34], we are developing an intelligent computer tutoring system called the **MENDEL** system. This system will help students become more knowledgeable problem-solvers.

In this paper, we will describe the logic of the **MENDEL** system as it generates genetics problems and offers tutorial advice to students. The MENDEL system is an example of the design approach to science education [35] because it encourages students to develop their understanding of genetics while they conduct experiments and test their hypotheses about genetics mechanisms against the resulting data. This calls for a student to entertain multiple hypotheses, tentatively treat each hypothesis as a conclusion, and construct a set of confirmatory/ disconfirmatory and logical/empirical arguments in support of the final conclusion. The tutorial component stays true to the design flavor of the open-ended problem-solving activity.

Finally the paper ends with a discussion of several larger issues that are involved in the design approach to science education: problem-solving with understanding; problem-based, experiential learning; the integration of rule-based with model-based reasoning; and, the role of human collaboration in machine-mediated learning environments. The **MENDEL** system described in this paper can be viewed as an experiment in applying the theoretical positions on learning, problem-solving and teaching to the design and use of computer software in education.

A Description of the MENDEL System

The MENDEL *system's goals*

The primary goal of the **MENDEL** system is to provide students with tutorial help to increase their conceptual understanding of genetics as well as their problem-solving skills. This is accomplished by creating a computer environment that will supplement (but not replace) laboratory problem-solving experiences in transmission genetics.

More specifically, the **MENDEL** system has the following goals:

1. to help students develop an understanding of genetics and genetics problem-solving. Students, in turn, will:
    a. improve their problem-solving performance,
    b. gain a better understanding of the conceptual structure of transmission genetics, and,
    c. improve their ability to explain and justify their problem-solving strategies in terms of the conceptual structure of genetics;
2. to help students develop their understanding of scientific research skills such as problem identification, hypothesis generation and testing,

data gathering and long-term inference making.

These two goals are intimately interconnected. They will be elaborated throughout the rest of the paper.

The MENDEL *system's components*

The **MENDEL** system has two primary components:

1. a problem GENERATOR program that includes:
    a. a CUSTOMIZE section, and,
    b. a problem-solving environment;
2. an expert TUTOR program that includes:
    a. a problem SOLVER,
    b. a problem-solving ADVISOR,
    c. a video/graphics LIBRARIAN, and,
    d. a student MODELER.

These components are summarized in Figure 1.

We have completed the GENERATOR program and a prototype of the problem SOLVER component. We are currently working on a prototype of the MODELER and ADVISOR components, and, are working on the design of the video/graphics LIBRARIAN.

Each of **MENDEL**'s components has a unique interface structure. The specific interfaces, however, are integrated into an overall visual interface on the IBM PC-AT screen. For example, each component embodies the following functions in a different way [16]:

1. reduce the working-memory load of a student;
2. aid conceptualization of the genetics content and problem-solving strategies;
3. decompose the problem into manageable subunits, and;
4. help structure the student's thinking.

The overall visual interface, on the other hand, tries to:

1. maintain a consistent command structure;
2. facilitate ease of interaction;
3. be visually-compelling and aesthetically pleasing;
4. be pedagogically sound with respect to the project goals.

The GENERATOR Program in the ***MENDEL*** *System*

The GENERATOR program is termed a "strategic simulation" and places students in a computer environment that simulates the problem-solving situations faced by transmission geneticists in a laboratory [28, 29].

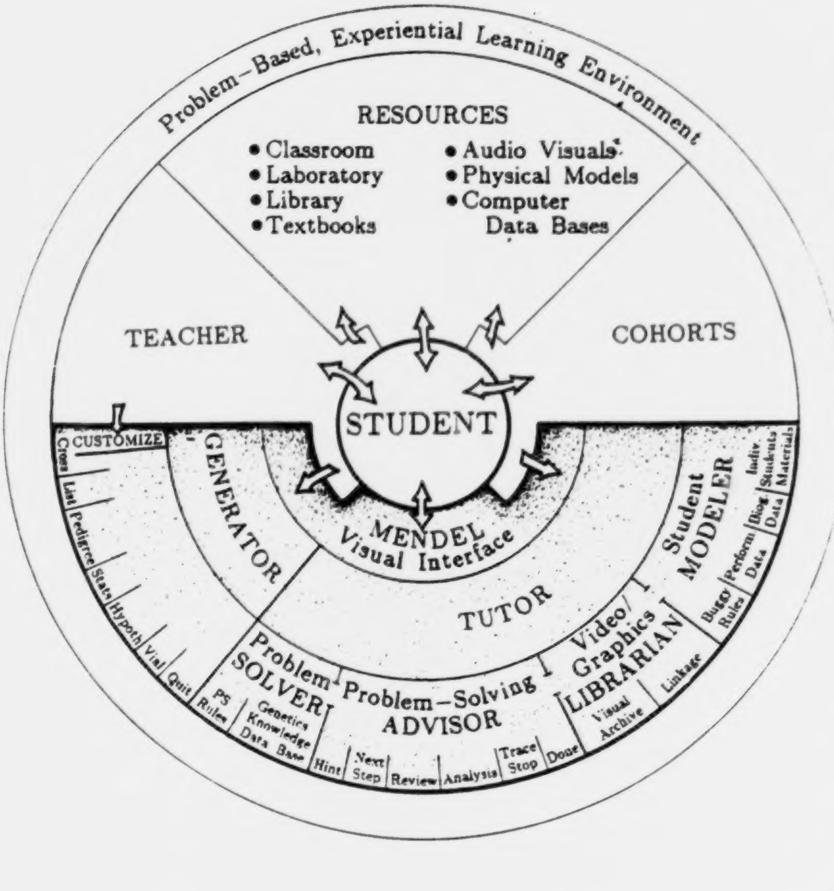

Problem-Based, Experiential Learning Environment
RESOURCES
• Classroom
• Laboratory
• Library
• Textbooks
• Audio Visuals
• Physical Models
• Computer Data Bases
TEACHER
COHORTS
STUDENT
CUSTOMIZE
Cross
List
Pedigree
Stats
Hypoth
Vial
Quit
GENERATOR
MENDEL Visual Interface
TUTOR
Student MODELER
Indiv. Students Materials
Biog. Data
Perform. Data
Buggy Rules
Video/ Graphics LIBRARIAN
Visual Archive
Linkage
Problem SOLVER
PS Rules
Genetics Knowledge Data Base
Problem-Solving ADVISOR
Hint
Next Step
Review
Analysis
Trace Stop
Done

Figure 1. Summary diagram of the **MENDEL** system's components.

Students who use the GENERATOR program have to pose their own problems and then use their genetics knowledge, their ability to perform genetics crosses, and their ability to use computational tools such as CHI square analysis to work out appropriate solutions. The students' experiences with the GENERATOR program are more realistic than those possible with textbook problems.

There are two parts to the GENERATOR program: a *CUSTOMIZE section* where users create classes of problems (within which cases are randomly generated later by the GENERATOR for students) and a *problem-solving environment* where users perform crosses to produce data and use data-management tools to manipulate and view the data (see Figure 1).

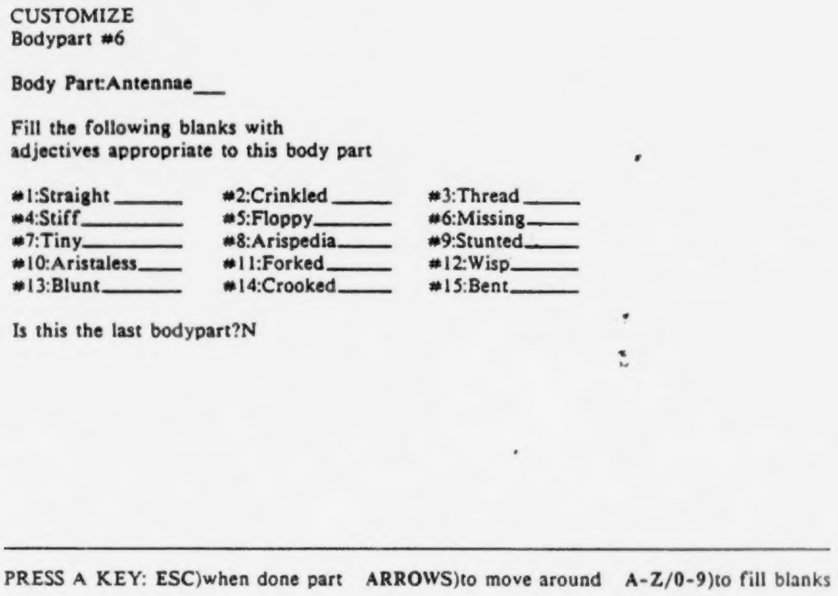
The CUSTOMIZE Section of the GENERATOR. Within the CUSTOMIZE section, a user can create *classes of problems* and define *sets of trait and variation names*. Classes of problems are created by filling in templates such as the one shown in Figure 2.

On each of these templates, the user can select the number (1-4) of traits for the problem, the range (1-99) of progeny from a cross and a set of *primary inheritance patterns:* simple dominance (the default value), codominance and multiple alleles. For each problem class, users can set the probability of the appearance of any particular inheritance pattern. In addition, users can select a set of *modifiers* to these primary inheritance patterns: sex linkage, lethality, penetrance, pleiotropy, gene interaction, and autosomal linkage. The modifiers can further be adjusted to set their maximum occurrence and probability of occurrence. For example, in the template shown in Figure 2, two inheritance patterns are possible in the same problem: simple dominance and codominance. Codominance, however, will never appear in more than one trait (since MaxCodom is set at 1) and it might not appear at all (since the CodomProb is set at 60%). These settings, as well as other genetics-specific parameters, permit a user to create a wide range of simple to very complex problems. Thus, the program can be used anywhere from junior high school up through graduate-level genetics.

Trait and variation names are also defined in the CUSTOMIZE section. A sample bodypart template screen for the Antennae trait is shown in Figure 3.

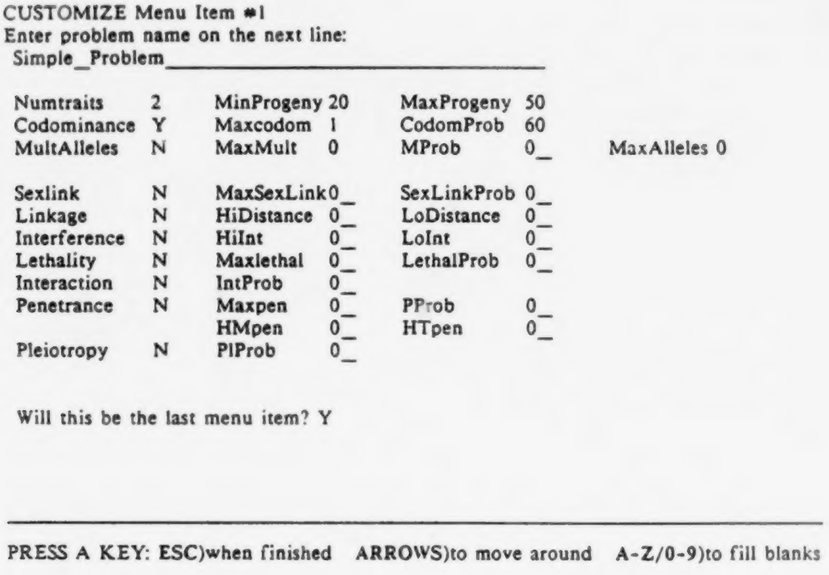
The traits (or Bodyparts) that might appear in any problem are selected along with variation names for that trait. In the sample problem to be discussed in this paper, we will use two body parts as traits: Antennae and Wings. The variables chosen in the CUSTOMIZE section of the GENERATOR "define" the problems that the user encounters in the problem-solving section.

The Problem-solving Environment of the GENERATOR. In the problem-solving section of the GENERATOR program, the student begins with a field-collected vial of organisms on the computer screen and then selects one of several functions. Figure 4 below depicts a "field-collected"



```
CUSTOMIZE
Bodypart #6

Body Part:Antennae___

Fill the following blanks with
adjectives appropriate to this body part

#1:Straight ______     #2:Crinkled ______     #3:Thread _____
#4:Stiff_________      #5:Floppy_________     #6:Missing_____
#7:Tiny__________      #8:Arispedia______     #9:Stunted_____
#10:Aristaless____     #11:Forked________     #12:Wisp_______
#13:Blunt________      #14:Crooked_______     #15:Bent_______

Is this the last bodypart?N

________________________________________________________________________________

PRESS A KEY: ESC)when done part   ARROWS)to move around   A-Z/0-9)to fill blanks
```

Figure 2. Sample Menu from the CUSTOMIZE Problem-Definition Screen.



```
CUSTOMIZE Menu Item #1
Enter problem name on the next line:
 Simple_Problem__________________________________

 Numtraits      2    MinProgeny 20    MaxProgeny  50
 Codominance    Y    Maxcodom   1     CodomProb   60
 MultAlleles    N    MaxMult    0     MProb       0_       MaxAlleles 0

 Sexlink        N    MaxSexLink0_     SexLinkProb 0_
 Linkage        N    HiDistance 0_    LoDistance  0_
 Interference   N    HiInt      0_    LoInt       0_
 Lethality      N    Maxlethal  0_    LethalProb  0_
 Interaction    N    IntProb    0_
 Penetrance     N    Maxpen     0_    PProb       0_
                     HMpen      0_    HTpen       0_
 Pleiotropy     N    PlProb     0_

 Will this be the last menu item? Y

________________________________________________________________________________

PRESS A KEY: ESC)when finished   ARROWS)to move around   A-Z/0-9)to fill blanks
```

Figure 3. Sample Menu from the CUSTOMIZE Bodypart-Definition Screen.

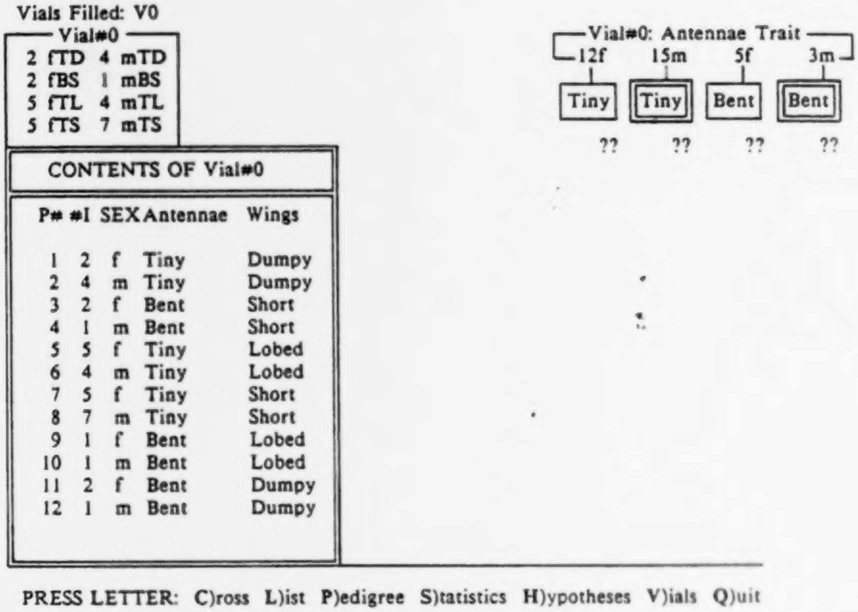
vial (i.e., Vial#0) whose contents have been elaborated by the List function.

Note that the vials on the computer screen display a shorthand representation of the trait's variation names (e.g., T = "Tiny"). A user can invoke the List option to see the full names of the traits and their variations. In addition, the graphic pedigree diagram on the computer screen represents a redescription of the Vial#0 data into a form that is appropriate for pedigree analysis. In this example, there are 12 females with tiny antennae (i.e., 2 Tiny/Dumpy, 5 Tiny/Lobed, and 5 Tiny/Short). The second variation names (i.e., Dumpy, Lobed and Short) refer to the Wings trait.

Figure 4 also shows some of the functions that are available to students:

C)ross	enables a student to cross individuals and obtain offspring;
L)ist	described above;
P)edigree	represents the vial data in a graphic form and is used by the problem solver to analyze the data produced from a cross experiment. The pedigree diagram is a useful, abstract redescription of cross data that makes it easier to see patterns and thus make inferences about genotypes across generations. The user's hypothesis about genotypes are entered over the question marks (underneath each pedigree box on the screen);
S)tatistics	allows the student to do mathematical calculations and CHI square tests with probabilities;
H)ypotheses	whereas the Pedigree option allows users to make specific hypotheses about parents and offsprings, the Hypotheses command allows users to enter hypotheses about the genetics of the population as a whole;
V)ial-options	helps students store and retrieve vials on the screen (for more space on the screen);
Q)uit	allows the student to abandon the current problem before going on.

Students who use the GENERATOR program are faced with an open-ended problem--how to explain the genetic mechanisms responsible for the phenotypes (i.e., appearance) of the population of organisms that



```
Vials Filled: V0
 Vial#0                                 Vial#0: Antennae Trait
2 fTD  4 mTD                             12f    15m    5f     3m
2 fBS  1 mBS                            Tiny   Tiny   Bent   Bent
5 fTL  4 mTL
5 fTS  7 mTS                               ??     ??     ??     ??

  CONTENTS OF Vial#0

P# #I SEX Antennae  Wings

 1  2  f  Tiny      Dumpy
 2  4  m  Tiny      Dumpy
 3  2  f  Bent      Short
 4  1  m  Bent      Short
 5  5  f  Tiny      Lobed
 6  4  m  Tiny      Lobed
 7  5  f  Tiny      Short
 8  7  m  Tiny      Short
 9  1  f  Bent      Lobed
10  1  m  Bent      Lobed
11  2  f  Bent      Dumpy
12  1  m  Bent      Dumpy

PRESS LETTER: C)ross L)ist P)edigree S)tatistics H)ypotheses V)ials Q)uit
```

Figure 4. Sample GENERATOR Screen of a Two-Trait Problem with the L)ist Option for Vial#0 (the Parental Vial).

they see on the screen. Underlying the generation of the field-collected vial and all subsequent offspring vials is a model of the inheritance patterns and modifiers as defined in the CUSTOMIZE component of the GENERATOR.

Within the context of the general problem, students are responsible for posing their own specific problems and for selecting the most appropriate approaches to a solution. This is done by performing crosses on the original set of organisms and/or successive generations and by doing statistical analyses. Thus, decisions such as whether enough data has been collected or what the results of statistical tests may mean must be made by students as they develop genetics-specific problem-solving strategies as well as more general scientific inquiry skills.

As rich as the GENERATOR environment is, it does not completely simulate the genetics laboratory experience. Aside from not having to feed, house, and mate actual organisms, students are also not faced with a critical first step in real genetics problem-solving--how to perceptually divide an organism into discrete, analyzable traits. This is already done by the GENERATOR program. Students therefore bypass the initial abstraction processes (of recognition and identification of traits and variations) involved in confronting data in scientific inquiry. In addition, they do not see many of the complex interactions that an organism's genotype (i.e., genetic makeup) has with its environment (both external and internal). These interactions can lead to a wide variation in the phenotype and are only approximated in the GENERATOR's environment. Nonetheless, GENERATOR- created experiences are far richer than the problem-solving experiences in typical undergraduate courses) [29].

The TUTOR program in the **MENDEL** *system*

The development of the TUTOR program has emerged from a consideration of the roles and responsibilities of a human tutor who is working with students in the GENERATOR environment. For example, a *human* tutor must be able to:

1. make inferences about the data generated by the student problem-solver;
2. maintain a history of a student's actions (including the crosses performed and the statements made about the data and crosses);
3. make inferences about the reasons for the student's problem-solving actions. These are drawn from a combination of what the student has done and has said. In so doing, the human tutor is building a model or representation of each student's or group of students' knowledge of genetics problem-solving;
4. compare the model of a student's knowledge with the tutor's understanding of the problem;
5. make decisions on the form of tutorial advice and the timing of this advice;
6. evaluate whether or not the student has benefitted from the advice.

Our work on the TUTOR component of the **MENDEL** system is guided by, but not necessarily limited to, these roles of a human tutor. Hence, we are developing a computer TUTOR that will be able to:

1. solve genetics problems;
2. interpret data generated by students;
3. develop a model of student knowledge;
4. compare this model with the TUTOR's knowledge;
5. decide whether or not to intervene;
6. decide on the nature of the tutorial intervention;
7. evaluate the success of the tutorial help.

In addition, our TUTOR will provide students with:

1. a set of computational tools for genetics problem-solving (Punnett squares, expression charts, etc.);
2. data-management tools to manipulate the data that they generate (pop-up calculators, data storage and retrieval, etc.);
3. graphical representation of genetics data and conceptual relations (pedigree and chromosome diagrams);
4. multiple windows into the reasoning of the TUTOR.

These last four features are normally *not* available from a human tutor.

The SOLVER Component of the TUTOR. In this section, we will present a simple example from its first appearance on the computer screen to a point where the inheritance pattern of one of the traits has been identified by the SOLVER. This will illustrate the internal logic of the SOLVER insofar as solving a problem is concerned although it will not indicate any tutorial interventions that might occur. This is an example of the TUTOR's TRACE-STOP mode of operation and will only be seen by students when they ask the ADVISOR within the TUTOR to solve an entire problem and explain its actions each step of the way. Because of the stochastic manner in which data is produced by the GENERATOR, two different TRACE-STOPs for the same problem would not be the same.

We begin with the GENERATOR-created screen of a two-trait problem shown in Figure 4. The goal is to infer which inheritance patterns and modifiers account for the distribution of phenotypic data in the population. Several actions can accomplish this goal: generating an hypothesis about a possible inheritance pattern and modifier, generating new data (i.e., invoking the GENERATOR program to perform a cross), checking to see if the data are consistent with the tentative hypothesis, and disconfirming alternate hypotheses. The TUTOR can perform each of these steps on its own because it has a SOLVER component that contains a high-level problem-solving Agenda and specific production rules for solving problems (see Figure 5 below for the SOLVER's Agenda).

This Agenda and related rules were extracted from research on how experts solve similar problems [26] and were formalized as condition/action relations (i.e., IF/THEN production rules). The SOLVER's Agenda items are described below along with a discussion of the example:

1. Redescribe Data from Initial Population for Each Trait: The first step in the Agenda directs the SOLVER to go to the GENERATOR-created population of organisms (see Vial#0 in Figure 4), extract key information (e.g., names and numbers of traits and variations) and store this information in the TUTOR's own internal data structures. It also directs the SOLVER to carry out some simple inferences that can be made from the initial population. For example, by focusing on the first trait (i.e., Antennae), the SOLVER can conclude that there are 12 female organisms and 15 male organisms with tiny Antennae in the initial population. Another example would be that the Antennae trait had only 2 variations (i.e., tiny and bent).

2. Entertain an Hypothesis about Inheritance Pattern: The redescribed data now serves as a set of "conditions" for the Solver's condition/action rules. Hence, the Agenda directs the SOLVER to search through its Hypothesis-Generating Rules (HGR) which in turn "fires" the following rule:

SOLVER AGENDA

1. **Redescribe Data from Initial Population for Each Trait**

2. **Entertain an Hypothesis about Inheritance Pattern**
   (hypothesis generation rules: HGR)

3. **Test Inheritance Pattern Hypothesis:**
   (find genotype to phenotype mapping)

   a. Make a cross (cross rules: CR)
   b. Redescribe data from a cross
   c. Explain cross in light of hypothesis
      (cross explanation rules: CER)
   d. Done?
      - If there are no consistent explanations, goto 2
      - If there is more than one explanations, goto 3
      - If there is exactly one explanations, goto 4
      - If there is absolutely no explanation, goto 1

4. **Check Your Result:**

   a. Make a prediction to test your hypothesis
   b. Are the crosses already performed consistent?
      (definitive cross rules: DCR)
   c. Disconfirm competing hypothesis
      (disconfirmation rules: DR)

Figure 5. Problem-Solving AGENDA for the SOLVER Component of the TUTOR.

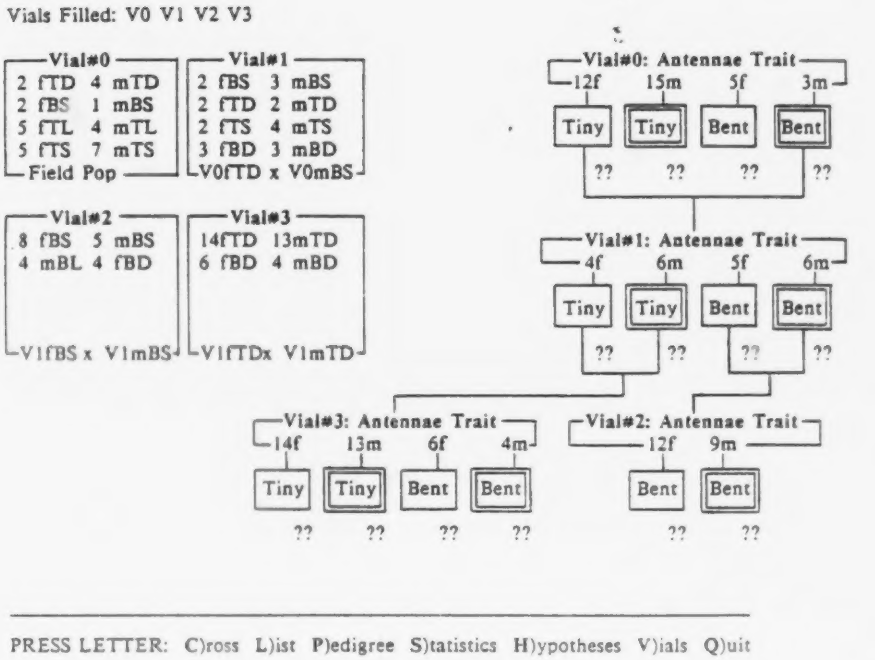
HGR1: **IF** (1) goal: generate an inheritance pattern hypothesis
(2) there are 2 variations for a trait
THEN assume simple dominance is the inheritance pattern for that trait

HGR1 states that after having broken the larger problem into a subproblem (i.e., focusing on one trait at a time), the SOLVER should proceed on the assumption that simple dominance may be the inheritance pattern responsible for the phenotypic data. This accomplishes several things. First, it simplifies the search space of possible underlying mechanisms that might account for the phenotypic data. Second, it makes a best first guess at such a mechanism the way an expert problem-solver would do. (Of course, there are several levels of genetics knowledge compiled into HGR1 which would have to be explained to a student who wanted to understand *why* this particular rule was a useful first guess). And finally, it translates a problem-solving strategy into a specific procedure. The SOLVER now has a way to match the phenotypic-level data against genotypic-level causal relationships.

3. Test Inheritance Pattern Hypothesis: The Agenda now directs the SOLVER to cross a female and male organisms from Vial#0. A Cross Rule (CR) fires because the appropriate conditions exist in the redescribed data. This rule directs the GENERATOR program to cross unlike variations (i.e., a tiny-antennaed female with a bent-antennaed male) because such a cross produces the most knowledge about the current hypothesis. (As mentioned above for rule HGR1, Cross Rules contain several levels of genetics knowledge). Hence:

CR2: **IF** (1) goal: plan a cross within a trait
(2) there is a variation, V1, for which you don't have a genotype
THEN cross unlikes: V1 with some other variation

The SOLVER also tells the GENERATOR to randomly choose one of the 12 female tiny-antennaed organisms and one of the 3 male bent-antennaed organisms. The resulting offsprings are placed in Vial#1. Figure 6 shows the computer screen at the end of the problem-solving session. For the time being, we need only focus on Vial#0 and Vial#1.



```
Vials Filled: V0 V1 V2 V3

--Vial#0--        --Vial#1--
2 fTD  4 mTD      2 fBS  3 mBS
2 fBS  1 mBS      2 fTD  2 mTD
5 fTL  4 mTL      2 fTS  4 mTS
5 fTS  7 mTS      3 fBD  3 mBD
--Field Pop--     -V0fTD x V0mBS-

--Vial#2--        --Vial#3--
8 fBS  5 mBS      14fTD 13mTD
4 mBL  4 fBD      6 fBD  4 mBD

-V1fBS x V1mBS-   -V1fTDx V1mTD-
```

Vial#0: Antennae Trait
12f
15m
5f
3m
Tiny
Tiny
Bent
Bent
??
??
??
??
Vial#1: Antennae Trait
4f
6m
5f
6m
Tiny
Tiny
Bent
Bent
??
??
??
??
Vial#3: Antennae Trait
14f
13m
6f
4m
Tiny
Tiny
Bent
Bent
??
??
??
??
Vial#2: Antennae Trait
12f
9m
Bent
Bent
??
??
PRESS LETTER: C)ross L)ist P)edigree S)tatistics H)ypotheses V)ials Q)uit

Figure 6. Sample GENERATOR Screen of a Two-Trait Problem Solved for the Antennae Trait.

The data in Vial#1 represent a new set of conditions for the SOLVER's rules to consider. Following the Agenda (see Figure 5), the SOLVER first redescribes the new data (Agenda item 3b) and then applies a series of Cross Explanation Rules (CER) (Agenda item 3c). One rule fires because the appropriate conditions in Vial#0 *and* Vial#1 exist. Hence:

CER6: **IF** (1) goal: explain a cross within a trait
(2) assumed inheritance pattern is simple dominance for that trait
(3) parents are of different variations
(4) offspring are of both variations
THEN (1) one parent and the offspring of the same variation are homozygous recessive
(2) the other parent and the offspring with this variation are heterozygous dominant

That is, the SOLVER finds that "unlikes" in the parents (tiny-antennaed and bent-antennaed) have produced "unlikes" in the offspring. If simple dominance was in fact the underlying mechanism in our example, the cross could be explained by the abstract genotypic pattern:

$$Aa \ x \ aa \rightarrow 1/2^{Aa} + 1/2^{aa}$$

The capital "A" in the genotypic pattern above represents the dominant allele and the lower-case "a" represents the recessive allele. The "Aa" represents a heterozygous allele-pair and "aa" a homozygous recessive allele-pair. Figure 7 summarizes all of the possible genotype-to-phenotype matches for the simple dominance case.

Of course, the SOLVER cannot at this point determine which specific genotype (i.e., Aa or aa) corresponds with which phenotype (i.e., tiny-antennaed or bent-antennaed) in Vial#1. The SOLVER therefore has to perform more crosses to establish such a correspondence.

At this point, the SOLVER continues to test the current inheritance pattern hypothesis (Agenda item 3d) because Vial#1 has added new conditions for the original set of Cross Rules. Hence, the following Cross Rule fires:

CR16: **IF** (1) goal: identify which of the offspring of an unlike cross are heterozygotes
(2) there are two variations in that offspring;
THEN consider crossing likes from this offspring.

The SOLVER therefore crosses two organisms of the same variation (i.e., bent-antennaed) from Vial#1. The results of the GENERATOR-created data are stored in Vial#2 (See Figure 6). Note that

Genotypic Level	Phenotypic Level	Cross Types	Number of Offspring Classes
1. AA x AA -> AA*	V1 x V1 -> V1**	likes	1
2. aa x aa -> aa	"	"	"
3. AA x Aa -> $_{1/2}$AA + $_{1/2}$Aa	"	"	"
4. Aa x Aa -> $_{1/4}$AA + $_{1/2}$Aa + $_{1/4}$aa	V1 x V1 -> $_{3/4}$V1 + $_{1/4}$V2	likes	2
5. AA x aa -> Aa	V1 x V2 -> V1	unlikes	1
6. Aa x aa -> $_{1/2}$Aa + $_{1/2}$aa	V1 x V2 -> $_{1/2}$V1 + $_{1/2}$V2	unlikes	2

*"A" represents the dominant allele, "a" the recessive allele.
"AA" represents the homozygous dominant allele-pair.
"aa" represents the homozygous recessive allele-pair.
"Aa" represents the heterozygous allele-pair.

**"V1" represents the first arbitrary variation, Notice that several genotypic patterns can underlie the same phenotypic pattern.

Figure 7. Relationship of Genotypic to Phenotypic Data for a Simple Dominance Case of Two Variations (V1 and V2) of One Trait (All Possibilities are Shown).

the SOLVER is now reasoning about the data from several generations of data. This strategy was chosen because it approximates optimal problem-solving performance--something that was not always displayed by the experts [26]. The SOLVER now redescribes the data in Vial#2 and tries to explain the data in light of the simple dominance hypothesis. A Cross Explanation rule fires because the SOLVER has found the correct conditions in both Vial#1 and Vial#2. Hence:

CER7: **IF**
(1) goal: explain a cross within a trait
(2) assumed inheritance pattern for that trait is simple dominance
(3) parents have like variations within this trait.
(4) parents are either heterozygous or homozygous- recessive
(5) offspring have the same variation within this trait as the parents
THEN parents are very likely homozygous-recessive while offspring are also very likely homozygous-recessive

CER7 helps the SOLVER conclude that the bent variation of the Antennae trait in Vial#2 is due to a homozygous recessive allele-pair. The reasoning proceeds as follows: the SOLVER has already established from the previous cross that the tiny-antennaed and bent-antennaed variations in Vial#1 are *not* due to a homozygous dominant genotype (i.e., the genotypic pattern

$$Aa \times aa \rightarrow 1/2^{Aa} + 1/2^{aa}$$

accounted for the data--thus excluding AA). Of the three simple dominance mechanisms that could account for the appearance of a bent-antennaed phenotype data in Vial#2:

$$AA \times AA \rightarrow AA$$

$$AA \times Aa \rightarrow 1/2^{AA} + 1/2^{Aa} \text{ (both \textit{appear} the same)}$$

$$aa \times aa \rightarrow aa$$

the first and second genotype patterns can be eliminated because both involve a homozygous dominant genotype. This leaves the homozygous recessive genotype pattern (i.e., aa x aa -> aa) to account for the data in Vial#2. By inference, the SOLVER can also conclude that the tiny-antennaed variation in Vial#0 is due to a heterozygous allele-pair (Aa) because that was the only other pair left in Vial#1. (The SOLVER fills in these hypotheses in the pedigree diagram in place of the question marks below the pedigree boxes on the screen for the benefit of the student.) At this point, the problem seems to be solved. However, there is one more step in the Agenda.

4. Check Your Result: The SOLVER has accounted for both variations of the Antennae trait in Vial#1 on the assumption that simple dominance was the case. The Agenda therefore directs the SOLVER to carry out one more step: checking the SOLVER's conclusion with an *independent* cross. Collins [26] has found that expert geneticists add a definitive cross of two heterozygous individuals at this point in the process. Hence, the SOLVER applies its Definitive Cross Rules (DCR) and fires the following rule:

DCR1:	**IF**	(1)	goal: become more confident in an inheritance pattern for a trait
		(2)	assumed inheritance pattern is simple dominance with a high degree of confidence
		(3)	heterozygotes have been identified
	THEN		cross the heterozygous individuals

This rule takes a previously-identified heterozygous individual from Vial#1 (i.e., tiny-antennaed), crosses a male and a female with this variation, and places the results in Vial#3 (See Figure 6). Again, because new data has been generated, new conditions exist for the application of the Cross Explanation Rules. This time, CER8 fires:

CER8:	**IF**	(1)	goal: explain a cross within a trait
		(2)	assumed inheritance pattern for that trait is simple dominance
		(3)	parents are heterozygous within this trait
		(4)	both traits are present within the offspring
		(5)	test comparing the ratios of offspring variations to 3:1 is significant
	THEN	(1)	increase confidence in identity of parents as heterozygous
		(2)	increase confidence in simple dominance as the inheritance pattern
		(3)	increase confidence that the parent's variation is dominant

This rule confirms that the tiny variation of the Antennae trait could only have come from a heterozygous allele-pair because only one simple dominance rule could account for this data:

$$Aa \times Aa \rightarrow 1/4^{AA} + 1/2^{Aa} + 1/4^{aa}$$

Notice that both AA and Aa show up as the same phenotypic variation in the offspring because the allele "A" is dominant to the recessive allele "a". Hence, a 3 to 1 ratio for phenotype characteristics is expected to show up in the offsprings (i.e., $3/4^{A-} + 1/4^{aa}$).

Notice also that, although we have confirmed the simple dominance hypothesis for this set of data, there still exists the slightest possibility that some other inheritance pattern and/or modifiers could account for the data. Most genetics experts in such a situation eliminate (or disconfirm) these possibilities with some standard disconfirming crosses[26]. Hence, the Agenda (Item 4c) directs the SOLVER to try out some final Disconfirming Rules (DR) such as:

DR1: **IF** (1) goal: disconfirm alternate hypotheses
(2) inheritance pattern is simple dominance
(3) sex-linkage is modifier under consideration
(4) a cross of a dominant male with a recessive female results in offsprings that are not limited to dominant females and recessive males
THEN sex-linkage modifier is not operating

The example discussed above illustrates the SOLVER's rule-based approach to generating hypotheses about inheritance patterns and to generating crosses within the constraints of these hypotheses. The example shows how rules are used for confirming and disconfirming hypotheses based upon the phenotypic data that emerge after each new cross. The SOLVER therefore has the ability to keep track of its own inferences and the ability to build up genetics knowledge appropriate to a given population of organisms. The TUTOR will have access to all of this information and can use it to provide tutorial advice.

Finally, the SOLVER, when solving problems on its own, performs all aspects of problem-solving. However, in the typical case, the SOLVER will not be making crosses. Rather, it will be suggesting crosses in light of certain student-chosen hypotheses and making inferences from student-generated data. In the latter case, the SOLVER works with the crosses that the student has made and then tries to extract as much knowledge as possible from this data in light of hypotheses that the *student* is entertaining.

The ADVISOR Component of the TUTOR: In the section above on the SOLVER, we described the user-requested TRACE-STOP mode of the ADVISOR. In addition to the TRACE-STOP mode, we will provide the student with other tutorial aids: HINT, NEXT-STEP, REVIEW, and ANALYSIS. Each of these commands can be categorized on two dimensions: one dimension deals with suggestions about a future action (HINT and NEXT-STEP) or an evaluation of past actions (REVIEW and ANALYSIS); the other dimension deals with specific actions (NEXT-STEP and ANALYSIS) or general strategies (HINT and REVIEW). These relationships are shown in Figure 8.

	General Advice (series of actions)	**Specific Advice** (single action)
Future Actions (SOLVER Data & Hypothesis)	HINT	NEXT-STEP
Past Actions (Student Data & Hypotheses)	REVIEW	ANALYSIS

[*] Other ADVISOR commands include the TRACE-STOP and DONE options.

Figure 8. User-Requested Tutorial Options of the ADVISOR Component of the TUTOR (Other ADVISOR commands include the TRACE-STOP and DONE options.)

Although we feel it is important for the ADVISOR to have the ability to decide when it is appropriate to offer advice (i.e., to have some TUTOR-initiated intervention strategy), we are currently focussing on what that advice will be. We have made a deliberate decision to implement the user-initiated advice-giving capabilities of the ADVISOR prior to and independently from the intervention strategy. This approach has many advantages. First, by having the student decide when he or she would like advice, we can have a workable tutor before actually implementing a TUTOR-initiated intervention strategy. Second, it is easier to add a more sophisticated intervention strategy to an existing advice-giving capability than it is to design both features at the same time. Finally, by implementing these capabilities independently, we can study the effectiveness of alternative intervention strategies (i.e., user-initiated vs. mixed-initiative interventions) before implementing any one.

We will now describe the user-initiated advice-giving capabilities of the ADVISOR:

1. The HINT Command of the ADVISOR: Students invoke the HINT option when they want a suggestion for what to do next. The ADVISOR then gives them general prompts, and, if that advice is not helpful, gives them increasingly specific hints. Even though HINT provides suggestions about future actions, these suggestions may make little sense to a student if there is something seriously wrong with what he or she has already done. In this case, the ADVISOR will comment on the error before providing a hint. If there is nothing seriously wrong, HINTs will be given that are appropriate to one of the following categories of action: performing crosses (via the Cross command); making hypotheses about individual or offspring class genotypes (via the Pedigree command); or making hypotheses about the genetics of the population as a whole (via the Hypotheses command). For example, if the SOLVER determines that it is possible to make a hypothesis about the genetics of the population, then the hints given to the student might proceed from general to specific as follows:

a. Hints to try to generate a hypothesis. For example: "Can you make any hypotheses? If so, please enter them."
b. Global redescription hints to help a student generate an inheritance pattern hypothesis. These include:
"What can you tell me about the initial population?"
    "How many traits? What are they?"
    "How many variations in each trait? What are they?"
"Have you done other problems with the same number of variations?"
    "What does the number of variations suggest to you?"
    "What if there were 3 variations instead of 2?"
c. Hypothesis generating hints (corresponding to HGR rules).

2. The NEXT-STEP Command of the ADVISOR: The NEXT-STEP command spells out exactly what the TUTOR's SOLVER would do next in

light of the student's current cross data and hypothesis. There are two possible next steps: *perform a cross* and *state an hypothesis.* When a student receives NEXT-STEP advice, he or she can ask why that advice was given by using the WHY command. In response to WHY, the rule that prompted the specific action is given. If the student seeks further explanation of this rule, the ADVISOR may offer [14],

a. *strategy explanations,* which the student requests by the CLARFY command, and
b. *support explanations,* which the student requests by the JUSTIFY command.

Strategy explanations are designed to clarify the rule by explaining it in terms of more general strategies applicable to many classes of genetics problems. Support explanations employ content knowledge and examples to justify the rule by describing or illustrating the genetic mechanisms underlying the rule.

For example, a student may have crossed Vial#0 individuals with the same phenotypes six times while indicating a current hypothesis of simple dominance. If the NEXT-STEP command is now invoked, the ADVISOR would recommend that the student use some of the offspring that have been produced and make a cross of individuals with unlike variations. If the student invokes the WHY command, the ADVISOR would present Cross Rule 2 (which was used earlier to illustrate the SOLVER's rules). If the student then invoked the CLARIFY command, the ADVISOR would offer a more general strategic explanation (e.g. that crossing unlikes makes it possible for a solver to either construct or identify heterozygous individuals). If the student still wasn't satisfied he or she could invoke CLARIFY again and get explanations of a more general nature, such as:

a. to match phenotypes with genotypes requires the identification of heterozygous individuals,
b. to test inheritance pattern hypotheses requires that all phenotypic variations be matched with genotypes, and,
c. one action in the solving strategy is to Test Inheritance Pattern Hypotheses (Figure 5, Agenda Item 3).

The purpose of CLARIFY is to help the student understand the specific advice provided by the NEXT-STEP command.

The student might also invoke the JUSTIFY command. CR2 relies on the empirical associations of the genotype-to-phenotype relationships illustrated in Figure 7. The tutor might justify crossing unlikes at this point in the problem-solving process by highlighting relationships 5 and 6--that when the variations of the parents are unlike, heterozygous offspring are produced. The next level of explanation would employ relationship 4 to illustrate how crossing parents with like variations can be used to match genotypes with phenotypes.

3. The REVIEW Command of the ADVISOR: The REVIEW command uses data from the student MODELER and possible student errors to look back over the student's performance and make appropriate comments. REVIEW is like ANALYSIS (described below) in that it looks back at student actions. However, REVIEW does a more general evaluation

based on student behaviors spanning the entire problem solution up to the point when a student asks for a REVIEW. REVIEW will make general comments about the student's strategy such as "You didn't use offspring as parents very often". Comments like this can be helpful to a student in future problem-solving sessions.

4. The ANALYSIS Command of the ADVISOR: Whereas the TRACE-STOP command walks students through a solution of crosses that were generated by the SOLVER, the ANALYSIS command walks students through the crosses that *they* made and points out what knowledge the SOLVER can extract from each cross. The ANALYSIS option then debriefs students about the potential significance that each cross had for the problem-solving process and where students may have made one or more of three types of errors: an inconsistent hypothesis, an unwarranted inference, or missed a warranted inference.

5. The DONE Command of the ADVISOR: The student invokes the DONE command when the problem is finished. The ADVISOR will then:

a. check the student's solution for consistency and point out inconsistencies,
b. check the student's solution for completeness and make comments about incompleteness,
c. allow the student to return to the problem-solving environment if they would like to continue working,
d. ask the student if they would like a REVIEW or an ANALYSIS.

The Video/Graphics LIBRARIAN Component of the TUTOR

The video/graphics LIBRARIAN manages both computer-generated graphics and visuals stored on a video disk. Each type of graphics information is accessible to the TUTOR when a decision has been made that a student would benefit from tutorial advice. The information in the video library will also be directly available to a student.

The graphics material will be invoked to provide support explanations (e.g. about meiotic events) to accompany tutorial advice. The graphics managed by the LIBRARIAN are of two types--fixed visuals from the video disk and interactive, computer-generated graphics. The fixed visuals will include, for example, both commercially-produced stills and moving visuals of actual cells undergoing meiosis as well as stylized equivalents that illustrate only the most salient features of meiosis. Such immediate access to high quality video materials is not typically part of genetics instruction.

The second type of visual materials under the management of the LIBRARIAN is computer-generated graphics. For example, an understanding of the mechanism of meiosis can help a student explain his or her solution to a problem (a desired learning outcome) and recognize trends in the data which may not correspond to a simple independent assortment pattern. Once students recognize such a situation, they can begin to think of how linkage (including variable map distances and/or interference) might help to explain the patterns observed in the data. We have chosen to work with meiosis first since it is so central to understanding genetics problem-solving and because students have difficulty understanding meiotic processes [21, 36]. One of the ways that we have done this is through the development of a module called LINKAGE.

When LINKAGE is invoked by the LIBRARIAN or the student, it can help the student better understand meiosis by providing an opportunity to test various hypothesis that they may have to explain their data. By invoking LINKAGE, the student can create customized chromosome/gene models. This is done by allowing the students to:

1. create chromosome/gene arrangements for two parental organisms;
2. vary the map distances separating any linked genes and turn interference *on* or *off*;
3. observe the chromosomes that they have created undergo meiosis;
4. select the number of offspring to result from crossing two parents;
5. observe the offspring phenotype distribution that results from the cross;
6. change any of the above variables and observe how the offspring phenotype data is effected.

Thus a student working with a three-trait problem might begin with a model in which each individual had three pairs of homologous chromosomes (e.g. where the chromosomes assort independently and therefore are not linked). Two individuals could be identified as parents and that offspring phenotype distributions for a specified number of offspring in that generation could be observed. It would then be possible to construct a single pair of chromosomes so that all three genes are on the same chromosome pair (e.g. linked) and do the exact same thing that was just done for the unlinked situation. The student constructs as many alternative chromosome/gene arrangements as desired, thus having relatively immediate opportunities to observe how multiple chromosome/gene models lead to different patterns in the phenotypic data. The importance of programs like this, which the LIBRARIAN manages, is not only that they serve a tutorial function, but they provide a student with opportunities to work with multiple models of phenomena--something that is common in science, but less so in science instruction.

The Student MODELER Component of the TUTOR. In order for the TUTOR to intervene in the student's problem-solving process with tutorial advice, it must have access to information about that student. The function of the student MODELER is to gather such information, make inferences from it about the state of the students's knowledge (both strategic and conceptual), and make that information available to the TUTOR.

At the very least, the MODELER must keep a history of student actions such as: the vials(s) from which organisms are selected for crosses, the making and checking of hypotheses, the making of inferences about the genotypes of individuals or phenotype classes, and if and when students do statistical analyses. Some of this information will be directly available from a student's interactions with the basic GENERATOR program (the vials from which parents were taken) or by taking advantage of other GENERATOR functions (statistics or the Pedigree chart function).

Beyond this, the MODELER will need to recognize patterns in a set of individual actions and to make inferences about some student actions. For example, it is possible to recognize quickly that a student is taking all parental organisms from Vial#0. Although a problem could be solved by doing this, it is not an ideal approach because it does not acknowledge the importance of looking at data from within a lineage of several generations. It is therefore necessary to recognize when a student either misses a warranted inference or makes an unwarranted inference. This could be done directly by noticing when a student fails to enter genotype information on the pedigree chart or enters an unwarranted genotype. In order to recognize either student action, or lack of action, it is necessary to make comparisons with what action the SOLVER could make in response to the same data.

A student solving problems will execute a set of actions similar to the SOLVER'S agenda. These actions can be modeled as problem-solving rules. In addition, there should be conceptual knowledge (more than rules or empirical associations) which underlie the rules. This causal knowledge (e.g. of meiosis) is the basis for problem-solving with understanding and model-based reasoning. Both rule-based and model-based reasoning are ultimately important [37]. Rule-based reasoning is easier for the MODELER to process, however, so we plan to develop this capability of the MODELER first. The MODELER's ability to infer student conceptual knowledge will be added gradually, bolstered by our research on novice knowledge of genetics and how that knowledge relates to problem-solving actions.

Concluding Remarks

In this paper, we have described an on-going research and development project that will result in a unique genetics problem-solving environment. The environment both simulates a transmission genetics laboratory and provides computer-generated advice. It is intended to supplement undergraduate genetics education although it is flexible enough to be used in high-school biology or graduate courses.

The **MENDEL** system embodies certain values and commitments to science education that have guided us in our design choices and research questions. Our commitments can be categorized around the following themes:

1. problem-solving with understanding;
2. problem-based, experiential learning;
3. integration of rule-based and model-based reasoning, and;
4. collaborative, machine-mediated learning environments that

embody the foregoing themes.

Our commitment to the importance of problem-solving with understanding (as opposed to efficient problem-solving performance *per se*) is based on our own experience as science teachers, our research on problem-solving, and our critical analysis of the potential dangers of mindless learning in computer-based education.

The importance of problem-solving with understanding was driven home in one of our studies with high-school genetics students who were using the GENERATOR program. At one point, when a group of these students was having a particularly hard time with one of the computer-generated problems, the instructor inadvertently suggested what our research had shown to be a very powerful problem-solving rule. The students henceforth applied that rule to similar problems without thinking of the underlying genetics mechanisms. We had inadvertently created students who mindlessly followed rules. This is not to suggest that we are against rules or rule-following. Rather, we want rules to emerge in the minds (and behaviors) of our learners as a result of experience and understanding. A tutor must therefore do much more than reveal problem-solving rules. This brings up our second commitment.

Problem-based learning is emerging as an alternative approach within medical education [38] and experiential learning is already well established in organizational theory and business education. [39] We have learned from these traditions as well as from our work on strategic simulations that long-term inferencing is best learned through a series of experiments and associated problem-solving activities [28, 29].

In many ways, problem-based, experiential learning is nothing new because most scientists learn to do science in this way. However, most students who take introductory science courses do not become scientists and therefore do not have this experience. At most, they get a simplified, sanitized, rational-reconstruction of science from a text book while sitting in large lecture halls. This is not science but a rhetoric of conclusions.

What we are trying to do is to offer these students some experience at conducting genetics experiments, generating and testing hypotheses, and developing some understanding of genetics problem-solving. The **MENDEL** system is one way to make this feasible. We realize that some aspects of problem-based learning and experiential learning cannot be simulated in our environment. For example, we do not include the initial abstraction stages of identifying traits and variations of organisms. How important perceptual discernment and abstraction are for genetics understanding remains an open research question. Whether we could use, or would want to use, the videodisc to simulate these initial stages of doing science also remains to be seen. We have chosen to give the videodisc a different role in our project.

Our version of problem-based, experiential learning provides students with significant and realistic transmission genetics problems to solve. Our environment then provides students with computational tools, graphical representation of genetics concepts, and tutorial advice that encourage conceptualization about the underlying genetics mechanisms. It does so by letting students pose questions, make conjectures (i.e., enter hypotheses), and learn from their experience (i.e., perform crosses, use computational tools). Conceptualization here refers to both genetics-specific content and the nature of scientific inquiry. This brings us to our next commitment.

As mentioned earlier, students are quite willing to stop at the rule-following level of problem-solving. However, students are also able to understand the reasons behind problem-solving strategies. We, as educators, therefore have an obligation to help our students reach their full potential. In science education, this means reaching a certain level of scientific understanding and scientific inquiry. We try to achieve this within the constraints of the **MENDEL** system by helping students use model-based reasoning as well as rule-based reasoning. Rule-based reasoning is aided by the TRACE-STOP and NEXT-STEP commands where students are presented with the heuristic problem-solving rules that the SOLVER uses. These commands present rules in the exact problem-solving situation to which they apply. Thus, the student can actively engage in applying the rule. Model-based reasoning is aided by the JUSTIFY command as well as by the LIBRARIAN's routines. For instance, the LINKAGE module of the LIBRARIAN will be used to explain rules for generating and testing linkage hypotheses in model-based terms.

A key aspect of model-based reasoning is that the solution to a problem is actually the hypothesis in the mind of the student throughout the problem-solving process. Students therefore have to develop problem-solving strategies that exercise their critical and judgmental faculties and not just their technical abilities. Students also have to be sensitive to the data that emerge in their experiments. Model-based reasoning therefore becomes the link between theory-directed and data-directed problem-solving. Model-based reasoning can also be seen as the key to understanding the empirical associations of problem-solving rules.

Problem-solving with understanding, problem-based, experiential learning, and model-based reasoning do not occur in isolation. They are not merely individual psychological processes in the mind of the learner but are inherently social processes. We therefore believe that this type of learning requires collaboration with others. We try to structure our problem-solving environment and our tutorial advice so that collaboration between students and tutors can take place. Furthermore, we have made our simulation of a genetics laboratory complex enough so that robust experimentation can take place (i.e., the GENERATOR is *not* a toy universe) and so that heuristic approaches to solving problems can take precedence over algorithmic approaches (e.g., where multiple conceptualizations and mixed data-driven and theory-driven approaches can take place). This is fertile ground for collaboration.

Our final commitment deals with how we believe computers should be used in science education. We believe that computers should be used for strategic simulations in order to *supplement* science education. Strategic simulations remain a rational reconstruction of scientific experiments, no matter how complex they become, and so can never replace actual experimentation. We also believe that computer tutors should play an advisory rather than in a supervisory role. Computer tutoring is a new type of tutoring rather than a substitute for human tutorial engagement. Human tutoring still remains central for science education. Our final commitment therefore translates into a vision of the computer as a science teacher's assistant.

Acknowledgements

We gratefully acknowledge the support of the following individuals and organizations: Dean John Palmer and the University of Wisconsin-Madison School of Education; the IBM Corporation for the support that they ave given to UW-Madison's Project TROCHOS; the FIPSE (Fund for the Improvement of Postsecondary Education) program of the US Department of Education (Grant# G008301471 and Grant# G0085410.39); and the National Science Foundation (Grant# MDR-8470277).

Special thanks go to Judith Van Kirk and Gary Price for many helpful comments on earlier drafts of this paper.

References

1. National Academy of Science & National Academy of Engineering. *Science and Mathematics in the Schools: A Report of a Convocation,* Washington, DC, National Academy Press, 1982.

2. E. P. Volpe, "The shame of science education". *American Zoologist*, 24:433-443, 1984.

3. D. Sleeman and J. S. Brown (eds), *Intelligent Tutoring Systems*, New York: Academic Press, 1982.

4. W. J. Clancey and R. Letsinger, "NEOMYCIN: Reconfiguring a rule-based expert system for application to teaching," in W. J. Clancy and E. H. Shortliffe (eds.), *Readings in Medicial Artificial Intelligence; The First Decade*, Reading, Mass: Addison-Wesley, 1984.

5. M. McCloskey, A. Caramazza, and B. Green, "Circular motion in the absence of external forces: Naive beliefs about the motion of objects." *Science*, 210 (4474):1139-1141, 1980.

6. A. A. DiSessa, "Unlearning Aristotelian physics: A study on knowledge-based learning." *Cognitive Science*, 6(1): 37-75, 1982.

7. M. Hackling, and D. Treagust, "Research data necessary for meaningful review of grade ten high school genetics curricula."

Journal of Research in Science Teaching, 21:197-209, 1984.

8. M. Smith, M. and R. Good, "Problem-solving and classical genetics: successful vs. unsuccessful performance." *Journal of Research in Science Teaching*, 21(9): 895-912, 1984.

9. J. H. Larkin, J. McDermott, D. P. Simon, and H. A. Simon, "Expert and novice performance in solving physics problems." *Science*, 208:1335-1342, 1980.

10. F. Reif, "Understanding and teaching problem-solving in physics. Research on physics education." *Proceedings of the First International Workshop on Physics Education. Paris*, Editions du Centre National de la Reserche Scientifique, 1983.

11. P. Hewson, "A conceptual change approach to learning science." *European Journal of Science Education*, 3:383-396, 1979.

12. K. Strike, G. Posner, P. Hewson, and W. Gertzog, "Accommodation of a scientific conception: Toward a theory of conceptual change." *Science Education*, 66: 211-227, 1982.

13. J. S. Brown, and R. Burton, "Diagnostic models for procedural bugs in basic mathematical skills." *Cognitive Science*, 2(2): 155-192, 1978a.

14. W. J. Clancey, "The epistemology of a rule-based expert system--A framework for explanation." *Artificial Intelligence*, 20:215-251, 1983.

15. B. Woolf, and D. D. McDonald, "Building a computer tutor: Design issues." *Computer*, Sept., 61-73, 1984.

16. J. R. Anderson, C. F. Boyle, R. Farrell, and B. J. Reiser, "Cognitive principles in the design of computer tutors." *Proceedings of the Sixth Annual Conference of the Cognitive Science Society*. Boulder, CO., 1984.

17. L. Steels, and J. A. Campbell (Eds.), *Progress in Artificial Intelligence.*, New York: John Wiley & Sons, 1985.

18. J. Lipson, C. V. Bunderson, B. Baillo, J. B. Olsen, and K. M. Fisher, "Evaluation of an intelligent videodisc for developmental biology instruction." *International Journal of Machine-Mediated Learning*, 1, 1983.

19. R. Tennyson, "Artificial intelligence methods in computer-based instructional design." *Journal of Instructional Development*,

7(3):17- 22, 1984.

20. T. Munakata, "Knowledge-based systems for genetics." *International Journal of Man-Machine Studies*. 23: 551-561, 1985.

21. J. H. Stewart and M. Dale, "Solutions to genetics problems: Are they the same as correct answers?" *Australian Science Teachers*, 27(3):59-64, 1981.

22. J. H. Stewart, "Student problem-solving in high-school genetics." *Science Education*, 67(4): 523- 540, 1983.

23. N. Thompson, and J. H. Stewart, "Secondary school genetics instruction: Making problem-solving explicit and meaningful." *Journal of Biological Education*, 19(1): 53-62, 1985.

24. W. Albright, "Problem-solving strategies used by university students to solve realistic genetics problems." (Master's Thesis, in progress).

25. S. Slack, "Problem-solving strategies used by high-school students to solve realistic genetics problems." (Master's thesis, in progress).

26. A. Collins, "Strategic knowledge required for desired performance in solving transmission genetics problems." (Ph.D. dissertation, University of Wisconsin - Madison), 1986.

27. J. R. Jungck, and J. N. Calley, *GENETICS: Strategic Simulations in Mendelian Genetics*., Wentworth, N.H.: COMpress, 1986.

28. J. R. Jungck, "Strategic simulations: Experiencing research-like problem-solving in computer biology modules." In L.R. Meeth & D.S. Gregory (Eds.), *Dunedin, Florida Studies in Higher Education*. 133- 134, 1982.

29. J. R. Jungck, and J. N. Calley, "Strategic simulations and post-socratic pedagogy: Constructing computer software to develop long-term inference through experimental inquiry." *The American Biology Teacher*, 47(1):11-15, 1985.

30. M. J. Streibel, "Dialog design and instructional systems design for an intelligent videodisc system. " *Videodisc and Optical Disc*, 43(3): 216-219, 1984.

31. M. J. Streibel, "A critical analysis of the use of computers in education." Accepted for publication in the *Educational Com-*

munications and Technology Journal.

32. M. J. Coombs, (Ed.) *Developments in Expert Systems*. New York: Academic Press, 1984.

33. M. H. Richer, and W. J. Clancey, W.J, "GUIDON-WATCH: A graphic interface for viewing a knowledge-based system." *I.E.E.E. CG&A*, November, 1985.

34. D. R. Peachy, and G. C. McCalla, "Using planning techniques in intelligent tutoring systems." *International Journal of Man-Machine Studies*, 24: 77-98, 1986.

35. N. Peterson, J. Jungck, D. Sharpe and W. Finzer, "A design approach to science: Simulated Laboratories, learning via the construction of meaning." *Machine-Mediated Learning*, 2:1, 1986.

36. J. Stewart,and M. Dale, "High school students' models of chromosome/gene behavior." MENDEL Technical Report #2, 1986.

37. P. A. Koton, "Empirical and model-based reasoning in expert systems." *Proceedings of the 9th International Conference on Artificial Intelligence*, 1: 297-298. Aug. 18-23, Los Angeles, 1985.

38. H. S. Barrows, and R. M. Tamblyn, *Problem-based Learning: An Approach to Medical Education*. New York: Springer Publishing, 1980.

39. D. A. Kolb, *Experiential Learning*. Englewood CLiffs, NJ: Prentice- Hall. 1984.

APPENDIX B

MENDEL Research Report #1

MENDEL Research Report #1

A Description of the Strategic Knowledge of Experts Solving Realistic Genetics Problems

Angelo Collins
James Stewart
April, 1987

Work on Project MENDEL has been supported by the Dean of the School of Education - University of Wisconsin-Madison; the IBM Corporation; and the US Department of Education's FIPSE program (Grant #G008301471).

The authors' addresses are:

Angelo Collins
251 Bluemont Hall
Kansas State University
Manhatten, Kansas 66506

James Stewart
226 TEB
University of Wisconsin-Madison
Madison, Wisconsin 53706

A DESCRIPTION OF THE STRATEGIC KNOWLEDGE OF EXPERTS SOLVING REALISTIC GENETICS PROBLEMS

Introduction

If reports such as Science and Mathematics in the Schools: Report of a Convocation (National Academy of science, 1982) are any indication, problem solving is a topic of special concern among science educators. Concurrent with this interest is the problem-solving research of cognitive scientists that provides science educators with insights into the nature of problem solving and which holds promise for educational practice.

One research approach used by cognitive scientists has been to study the problem-solving performance of experts in content-rich domains, especially physics. In an early study, Bhaskar and Simon (1977), studying an expert in thermodynamics, noted the consistent use of a single problem-solving strategy, means/ends analysis. They also noted that the expert was consistent in performing a check of the solution. Chi, Feltovich, and Glaser (1981), comparing experts and novices solving mechanics problems, found that experts describe a problem in terms of the concepts of mechanics rather than in terms of incidental surface features. Larkin (Larkin & Rainhard, 1984; Larkin & Reif, 1979) claimed that physics experts begin solving a problem by constructing descriptions of the problem at several levels. These levels include a basic description taken from the facts of the problem statement, a scientific description which converts the facts to scientific concepts, and a computational description which reduces the relationships of the concepts to mathematical formulae. In a summary of their research on the problem solving performance of physics experts, Larkin, McDermott, Simon, and Simon, (1980) identified four characteristics of expert performance: 1) the conceptual knowledge of the expert is stored and retrieved hierarchically; 2) experts have ancillary knowledge of when and how to use the conceptual knowledge; 3) experts begin to solve a problem by redescribing the data given in the problem statement in conceptual terms and mathematical relationships; and 4) experts, solving typical problems, use a forward-working, knowledge-producing strategy such as setting subgoals.

In addition to the data gathered on problem-solving performance, the research on problem solving in physics demonstrated the power of non-statistical, small-n research. The power of this approach lies in the rich descriptions of subjects' problem-solving behavior that the researcher obtains using thinking aloud as a data gathering technique. The term thinking aloud is used to describe the solver's verbal reports of thoughts and images used while working on a problem. The thoughts are recorded and the verbatim transcripts are used by the researcher to infer the strategic knowledge of the solver.

Synthesizing much of the research in problem solving in physics and providing a framework for further research, Reif (1983a; 1983b) has designed a comprehensive model for understanding and teaching problem solving in any science discipline. The comprehensive model includes models of: desired performance derived from descriptions of expert performance, novice performance, learning and teaching. The two components of the performance models are the two types of knowledge required to solve problems, which Reif designates as *content knowledge* and *strategic knowledge*. He identifies three aspects of content knowledge; 1) the concepts and principles of the discipline; 2) the ancillary knowledge of when and how to use this content knowledge; and 3) the structure of this content knowledge. He also identifies three categories of strategic knowledge: 1) data redescription strategies which enable the solver to identify the essentials of a problem and limit the problem space; 2) solution synthesis strategies by which the solver plans and executes ways to search the problem space; and 3) solution assessment strategies by which the solver decides if the answer is as complete and accurate as possible.

Although physics was the first science discipline in which problem solving was studied, transmission genetics is receiving increased attention from researchers. Paralleling the research in physics, Smith & Good (1983, 1984a, 1984b) have described the strategies of experts solving genetics problems. They identified 32 tendencies that can be used to differentiate between expert (or successful) and novice (or unsuccessful) problem-solving performance in genetics. The tendencies of successful solvers that included: 1) that they perceive a problem as a task requiring analysis and reasoning; 2) that they use knowledge-producing (forward-working) strategies, 3) that they begin solving the problem by investing initial time in qualitatively redescribing the problem; 4) that they make frequent checks of their work; and 5) that they use accurate bookkeeping procedures. Smith and Good found that experts also have a fund of accurate genetics knowledge which includes models of procedures for problem solving.

The problems used by Smith and Good were challenging -- they required the solver to analyze data about offspring and infer the genetic causes of the data. But the problems were taken from textbooks, and textbook problems tend to require students to use relatively few, and recently-taught, concepts to obtain solutions. Textbook problems are well-structured whereas real problems in science tend to be ill-structured and require that the solver determine what conceptual knowledge is needed to obtain solutions. The performance of experts solving real problems has been studied in the field of medical diagnosis. Shulman, Elstein and Sprafka (1978) have identified several characteristics of medical diagnosticians who were judged by their peers to be highly successful. These characteristics include: 1) that they are not limited to the cues (data) in the original problem situation but continuously produce additional data; 2) that the strategy used most often to make a diagnosis (solve a problem) is hypothesis testing; 3) that expert diagnosticians entertain several hypotheses simultaneously; 4) and that hypotheses are confirmed, revised or discarded in light of additional data.

Computer simulations make it possible to create realistic problem-solving environments in which the problems are ill-structured, like real problems. Real problems in transmission genetics are not only ill-structured but also differ from typical textbook problems in form. In textbook problems, the solver is presented with a description of a trait (for example, height in pea plants) and variations (for example, tall and short) of parents, and the inheritance pattern (for example, simple dominance) controlling the production of offspring. Given the limited, static data, the solution is to predict the distribution of the variations among the offspring (3/4 of the offspring will be tall and 1/4 of the offspring will be short). To reach a solution requires cause-to-effect reasoning, that is, from the inheritance pattern to the distribution of variations among the offspring. In real genetics problems the researcher begins with observations about a population of organisms. The researcher selects parents with traits and variations of interest (decides what the problem is) and produces generations of offspring (data) until an inheritance pattern can be inferred. To reach the solution requires effect-to-cause reasoning. Realistic, computer-generated problems in genetics, such as problems generated by GENETICS CONSTRUCTION KIT (Jungck & Calley, 1984), provide an opportunity for students to learn to solve problems which lack structure.

Stewart (in press) claims that learning to solve realistic problems provides students with the greatest potential for achieving four important learning outcomes. These are: 1) knowledge of the concepts of a discipline; 2) the ability to recognize and use general problem-solving strategies; 3) the ability to apply general and discipline-specific problem-solving strategies; and 4) to understand aspects of the nature of science. In genetics, solving realistic problems provides students with opportunities to pose the problem, to use their knowledge of genetics to generate and evaluate data, and to arrive at justifiable explanations of their solutions.

A description of the strategic knowledge of experts solving realistic transmission genetics problems can contribute to the theoretical knowledge about problem solving in science by providing insights into the characteristics of successful problem-solving performance on realistic genetics problems. A description of the strategic knowledge of experts can also provide science educators with insight in designing instruction to enable students to learn to solve realistic problems.

The primary purpose of this report is to describe the problem-solving strategies of experts solving realistic, computer-generated, transmission genetics problems. A secondary purpose is to suggest implications for instruction in solving realistic genetics problems.

Methods

Jungck & Calley's, (1984) GENETICS CONSTRUCTION KIT (GCK) was the strategic simulation program used to generate realistic transmission genetics problems. The simulation begins by displaying a population of field-collected organisms with the sex and phenotype of each individual identified. The solver then selects individuals for parents and crosses them to produce offspring. Generations of offspring can be produced until the solver is able to infer the inheritance pattern operating on the population. Inheritance pattern is the term used to summarize the genetics knowledge required to match a phenotype (the trait and variation observed, for example, green pea pods) with the genotype (the abstract, theoretical genetic factors causing the variation, often a pair of alleles expresses as symbols such as 'Gg'). A problem must have an inheritance pattern for each trait and these inheritance patterns are mutually exclusive. The most common inheritance patterns taught in introductory biology are simple dominance, codominance, and multiple alleles. After the inheritance pattern has been inferred, the solver may decide that a modifier is also operating on the population. Modifier is the term used to describe a condition that may alter the distribution of phenotypes within an inheritance pattern without affecting the genotype-to-phenotype match. For example, the position of the alleles on the chromosome may result in some traits frequently being inherited together. Modifiers cannot exist independently of an inheritance pattern and more than one modifier may affect a single inheritance pattern at the same time. The modifiers usually taught in introductory biology include sex linkage and autosomal linkage.

GCK can be programmed to generate populations of many types of organisms. In this study the phenotypes of the organisms were traits and variations of insects. In a GCK problem an organism may have up to four traits. GCK organisms are diploid with homogametic females and heterogametic males. With GCK it is possible to construct problems with the following phenomena within the domain of classical Mendelian or transmission genetics: 1) simple dominance (dominance-recessiveness); 2) codominance; 3) sex linkage; 4) pleiotropy; 5) epistasis and other gene interactions; 6) lethality; 7) multiple alleles; 8) penetrance; 9) autosomal linkage; 10) multifactorial inheritance with and without environmental effects; and 11) complex combinations of most of the preceding phenomena (Jungck & Calley, 1986).

The parameters actually used to construct classes of problems in this study were: number of traits - two; inheritance pattern - simple dominance, codominance, or multiple alleles; modifier - six linkage or autosomal linkage. These classes of problems were chosen because they are typical of those used in high school and undergraduate biology instruction.

Seven experts solved realistic GCK-generated problems. All of the experts have doctoral degrees and experience in both teaching and doing research in genetics. Each expert spent an hour with the researcher learning the mechanics of the computer program. At this time the experts were given the list of phenomena possible for problems generated by GCK, but were not told the parameters actually used in constructing the problems they were about to solve. After the initial hour, in order to eliminate discomfort and/or silent clues possible if the researcher were present, each expert spent four additional hours alone solving problems. Because the experts worked at their own pace and because the problem generator was random, every class of problems was not addressed by every expert and some experts did more than one problem in a class. The classes of problems attempted by each expert are presented in Table 1.

In the initial session with the researcher, the experts were also asked to think aloud while solving the problems. They were given written directions on thinking aloud such as "Don't mumble". On the written directions were questions to ask themselves, such as "Why are you making the cross you are making?" with suggestions of points in the problem-solving process to remind themselves to think aloud, such

Table 1

Problems Attempted by Each Expert by Problem Class

EXPERT	1	2	3	4	5	6	7	TOTAL
PROBLEM								
Simple Dominance	2	2	2	2	2	2	2	14
Codominance	2	2	1	2	2	2		11
Multiple Alleles	4	1		1	1	2		9
Sex Linkage	1	1	2	1	1		1	7
Autosomal Linkage	2	1	1	1		1	1	7
TOTAL	11	7	6	7	6	6	4	48

as while the program is producing offspring from a cross. It was also emphasized that the transcripts of the tapes of them thinking aloud provide part of the raw data of educational research, and that too much data is preferable to too little data.

Two types of data were available for analysis: 1) the transcripts of the thinking aloud protocols and 2) the computer printouts of the sequence of crosses made by each expert for each problem, including the expert's solution. These data are termed research data to distinguish them from the data about offspring generated by the expert while solving the problem, which are termed problem data. A sample protocol and a sample printout for a problem are found in Figures 1 and 2 respectively. The class of problems from which the protocol and printout are taken is a two-trait problem with a simple dominant inheritance pattern and no modifiers. This problem and this class of problems will be used as examples in the analysis.

Analysis

The analysis and reduction of the data gathered from the performance of experts solving realistic genetics problems occurred in four stages. The first stage was to express the research data in terms of the concepts and principles of transmission genetics and group them into one of three categories: 1) about the problem data; 2) about an hypothesis about genotype to phenotype mappings that explains the results of a single cross, called a specific hypothesis; and 3) about an hypothesis about the inheritance pattern that could explain all the crosses and predict the results of additional crosses, called a general hypothesis. This first stage of data reduction required

Figure 1

Transcript of Think Aloud Protocol for Simple Dominant Problem

Well. fortunately we're back to 8 phenotypes and two groups of characteristics.

Yellow and straw and red and lobed.

Start with a dihybrid cross.

WE'll just for fun assume that the least frequent genotype, phenotype is going to be doubly recessive and do it.

That means it's SL. (straw and lobed)

I'll start with an SL by SL mating.

And we got all SL's.

That's helpful.

Let's try a YR by SL cross and then do an F(2).

If it works the way I'm expecting.

OK YR by SL gives uh only YR's.

So presumably I happened to pick up a homozygous YR and now I have just heterozygous YR's.

So we should get a nice distribution by crossing them.

Let's see if this new line is basically a 9:3:3:1.

20:9:5:2 which is very, very close.

So I'm sure I know what is going on already.

Might as well confirm it.

Doing a test cross.

Let's see Vial 2 by Vial 3.

That gives a 14:10:8:8 which I'm sure is near enough to 1:1:1:1.

Y and R are independently segregating and are dominant over S and L.

Figure 2

Computer Printout of Simple Dominant Problem and Solution

```
        Contents of Vial =1 (field collected population):
8 F  Yellow          Red                 7 M  Yellow          Red
1 F  Straw           Lobe                1 M  Straw           Lobe
3 F  Straw           Red                 1 M  Straw           Red
2 F  Yellow          Lobe                1 M  Yellow          Lobe

Entering CROSS....
     Vial #1         Phenotype #3        Individual #1  (f  SL x m SL)
     Vial #1         Phenotype #4        Individual #1

        Contents of Vial #2 (offspring from cross above):
16 F Straw           Lobe                11 M Straw           Lobe

Entering CROSS...
     Vial #1         Phenotype #1        Individual #2  (f YR x m SL)
     Vial #2         Phenotype #2        Individual #2

        Contents of Vial #3 (offspring from cross above):
20 F Yellow          Red                 28 M Yellow          Red

Entering CROSS....
     Vial #3         Phenotype #1        Individual #7  (f YR x m YR)

        Contents of Vial #4 (offspring from cross above):
10 R Yellow          Red                 10 M Yellow          Red
 3 F Yellow          Lobe                 2 M Yellow          Lobe
 1 F Straw           Lobe                 1 M Straw           Lobe
 2 F Straw           Red                  7 M Straw           Red

Entering CROSS....
     Vial #2         Phenotype #1        Individual #8  (f SL x m YR)
     Vial #3         Phenotype #2        Individual #5

        Contents of Vial #5 (offspring from cross above):
6 F  Straw           Red                 8 M  Straw           Red
6 F  Yellow          Red                 2 M  Yellow          Red
5 F  Yellow          Lobe                5 M  Yellow          Lobe
5 F  Straw           Lobe                3 M  Straw           Lobe

Solver's Solution
     Dihybrid.  Alleles Y and R are dominant over S and L, respec-
     tively.  They appear to be completely independently segregat-
     ing.
Correct Answer
     Trait #1 (Body):  There are 2 alleles.
     Genotypes map to phenotypes as follows:
          1,1  IS   Yellow     2,2  IS   Straw      1,2  IS   Yellow
     Trait #2 (Eyes):
     Genotypes map to phenotypes as follows:
          1,1  IS   Red        2,2  IS   Lobe       1,2  IS   Red
```

four steps, an example of which is shown in Table 2. Step 4 was to illustrate the dynamic, non-linear nature of the solution process.

The second stage in the reduction of the research data was to tabulate all the data refined in the first stage for all solvers for one class of problems. A table was constructed for each cross. Table 3 is the table for the first cross for all experts for the simple dominant problems they did. Comments about problem data are coded in the row labeled "redescription." If there was a comment on the number and types of variations, the code is "v". Comments on the number of classes of phenotypes are coded "c". Comments on missing classes of phenotypes are coded "m". If the expert used symbols such as letters instead of words to discuss the traits or variations, the symbol row is marked. For example, in Table 3, in the first column, the solver refers to the straw, lobed class of phenotypes as the "SL group". Comments about general hypotheses were coded. For example, SD is the code for simple dominant. To code the research data about the specific hypotheses, a chart was constructed of six possible crosses based on the phenotypic variations of the parents and the offspring produced. Each cross was assigned a letter which was used for coding. For example, specific hypothesis C is the cross of homozygous (individuals with like alleles, aa) recessive parent with another homozygous recessive parent producing offspring with one variation the same as the parents. Specific hypothesis F is the classic Mendelian cross of heterozygous (individuals with unlike alleles, Aa) parents producing offspring with two variations in a 3:1 ratio. The row labeled "type of cross" was a quick reference to the parents having the same variation

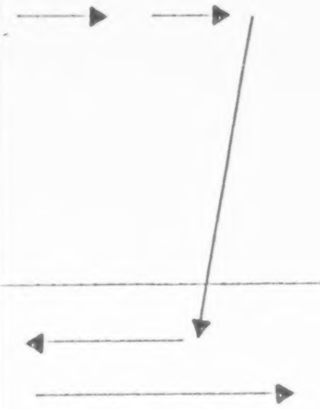
CROSS	Step 1 Read the transcript and mark it to correspond with the crosses	Step 2 Place the phrases of the transcript in groups depending on whether they refer to problem data (PD), specific hypothesis (SH), or general hypothesis (GH)	Step 3 Reduce the phrases of the transcript to transmission genetics concepts and add notes	Step 4 Draw arrows to represent the sequence and relationship of (PD), (SH), and (GH)
0	we're back to 8 phenotypes & 2 groups of characteristics yellow & straw & red & lobed. Start with a simple dihybrid cross. we'll just for fun assume that the least frequent phenotype is going to be doubly recessive & do it.	PD 8 pheno 2 group charac. yellow & straw red & lobed. SH least frequent is doubly rec. GH simple dihybrid	PD classes traits variations SH aa x aa aa double note lfp= rec GH Simple Dom	PD SH GH 
1	I'll start with an SL by SL mating & we got all SL's. That's helpful.	PD all SL SH SL x SL mating GH helpful	PD traits SH aabb x aabb aabb GH confirm SD	

Table 2 -- Stage One:
Data Reduction - Simple Dominance

Table 3 -- Stage Two:

DATA TABULATION - SIMPLE DOMINANCE

Cross 1

Person	1		2		3		4		5		6		7	
Problem	1	2	1	2	1	2	1	2	1	2	1	2	1	2
Redescription		m	v		c v not like real								m	
Symbol	x		x	x							x	x		
General Hypothesis	SD	other SD			SD		SD		SD		SD	SD		SD
Specific Hypothesis	C	C D			E		c		F		C	C		C/A D
Type of Cross	L	L U	U	U	U	L	L	U	U	L	L	L	L	L
NOTES		separate traits ratio sex ratio	series	separate traits sex ratio create het	separate traits ratio create h cross error	ratio sex ratio cross error		separate traits ratio create het create miss class	separate traits	have hypo but no confirm	separate traits missing class		sex ratio series	separate traits

(L for like) or different variations (U for unlike). Observations about the research data that were not easily coded were noted in abbreviated form in the last row.

In the third stage of analysis, the tabulated data were grouped into the three categories of strategic knowledge, to describe the performance of all the experts for each class of problems. Table 4 is the summary of the research data about problem data redescription for simple dominant problems; Table 5 is the summary of research data about hypothesis testing, the solution synthesis strategy used in simple dominant problems; and Table 6 is a summary about confirmation, the solution assessment strategy used in simple dominant problems.

The fourth stage of the analysis was to combine all the research data about the strategic knowledge of experts solving all the classes of problems considered in this study. The result of this analysis is the description of the strategic knowledge of experts solving realistic computer-generated transmission genetics problems which follows.

Data Redescription Experts tend to use data redescription is to isolate the essentials of the problem and limit the problem space. The experts include in their data redescription statements about the number and name of the traits and variations. They also combine individuals with the same phenotypic variations and consider classes of phenotypes. Identifying the number of variations for each trait and the number of classes of phenotypes is helpful in forming an hypothesis about the inheritance pattern. For example one expert begins:

> "So we have floppy vs. straight as a phenotype for something and expanded and vestigial about wings. We seem to have

Table 4

Data Redescription -- Simple Dominance

1. Details of Initial Redescription

--14 of 14 problems have some type of initial redescription

--10 include comments on traits, variations and classes of phenotypes

-- 2 include comments on traits and variations

-- 2 include comments on the number of classes of phenotypes

-- 5 note missing classes

-- 4 note least frequent phenotypes; of these, 1 also notes most frequent phenotype

2. Additional Occasions of Redescription

-- 2 problems are redescribed when the attention of the solver is focused on the second trait

-- 6 problems are redescribed whenever an alternate hypothesis is considered

-- 4 problems are resescribed at the end of the problem

Table 5
Solution Synthesis -- Simple Dominance

---

1. Origin of the General Hypothesis

   - 6 problems have the simple dominant inheritance pattern stated from the redescription of the initial population
   - 6 problems have hypothesis stated after 1 or 2 crosses
   - 2 problems have hypothesis stated after beginning a series of 4 or 5 possible crosses

---

2. Definitive Cross

   - In 8 of the 11 successfully solved problems a monohybrid or dihybrid F(2) cross is used to match genotype to phenotype
   - In 2 of these the heterozygote is constructed
   - In 6 an obligate heterozygote is located
   - In 3 of 11 successfully solved problems the linkage cross is used to match genotype to phenotype
   - In 3 an obligate heterozygote is used

---

3. Alternate Hypotheses

   - In 11 problems autosomal linkage as a modifier is considered and rejected
   - 11 times after the inheritance pattern is confirmed
     - 7 times by the linkage cross
     - 4 times by a dyhybrid F(2) cross

   - In 10 problems the sex linkage modifier is considered and rejected
     - 6 times after the inheritance pattern is confirmed
     - 2 times after the second cross
     - 2 times it is rejected by the sex linkage cross
     - 8 times the hypothesis is rejected because there is nothing to support it
   - In 1 problem lethality is rejected because there is nothing to suggest it
   - In 4 problems other hypotheses are considered -- sex influence, sex limited and interaction

Table 6

Solution Assessment - Simple Dominance

---

1. Mathematical

   - -- In 8 of the 8 problems that use an F(2), ratios are used to confirm the inheritance pattern and genotype to phenotype match
   - -- In 1 problem Chi square is used
   - -- In 7 problems the solver says the ratio "looks ok"
   - -- In 3 problems Chi squared is mentioned but not used

---

2. Strategic

   - -- In 6 problems both an F(2) and a linkage cross with an examination of their ratios are used to confirm simple dominance
   - -- In 4 problems the definitive cross is repeated with different individuals, in 1 case the reciprocals of the F(2) cross
   - -- In 9 of 11 problems at least two methods of confirmation are used

> two phenotypes for each of two characteristics in pairwise combinations. It could be, although we have no assurance of it, a simple case of two loci perhaps independent affecting two different characteristics."

In addition, the experts note any missing classes of phenotypes. For example, one expert says

> "...there are eleven different kinds, we've got eyes and bristles. There are only two types of bristles, hairless and singed, but for eyes we've got apricot, red, plum...Now what combination is not there...Let's count up...There are 1, 2, 3, 4, 5, kinds of females and 6 kinds of males. So we're missing a class of females."

A missing class of phenotypes by sex among the offspring of a cross may indicate that the sex linkage modifier is operating in that population. A missing class of phenotypes by variation or an unbalanced distribution of individuals by variation is an indicator that the autosomal linkage modifier might be operating in the population.

Data redescription always precedes the formulation of an hypothesis about an inheritance pattern or modifier. Therefore, for example, data redescription occurs at the beginning of the problem. One person begins

> "In this problem I suppose that all three genotypes are expressed as different phenotypes for tiny, specked and sable which would mean codominant or else that there are more than two alleles at the locus."

Experts also redescribe the problem data in the course of the solution synthesis whenever an alternate hypothesis is formulated. Alternate hypotheses are formulated 1) when a cross produces new data that alters the essentials of the problem; 2) when the solver is unable to infer or confirm an inheritance pattern; and 3) when solvers realize they have made an error in data interpretation. One example of new data altering the problem is:

> "Even before I begin I am suspicious that there is something funny because there are no b (blistery wing) males...I'll do a bs (blister wing, sepia eye) female with an ss (short wing, sepia eye) male cross...Oh, there are b (blistery wing) males, so much for that hypothesis. Now there are 8 groups and it looks like it is simple (dominance)."

Data redescription also occurs when a solver considers a hypothesis about a modifier and, in a multi-trait problem, when the solver begins to focus on the inheritance pattern of a different trait. In considering a modifier one expert says:

> "I crossed an sc (scarlet ocelli, crinkled antennae) by a wb (white ocelli, blunt antennae) and Wow, yeah I got - 2 wc's (white ocelli, crinkled antennae), 1 sb (scarlet ocelli, blunt antennae), 20 sc's (scarlet ocelli, crinkled antennae) and 11 wb's (white ocelli, blunt antennae). I can see clearly that I got an excess of parental types contributing to the heterozygotes that I used in the cross which suggests strongly that these are not independently assorting but linked."

By redescribing the data, the solver is able to limit the problem space to reasonable general hypotheses and consolidate and recall knowledge that has been obtained from the crosses that have been done so far.

Solution Synthesis Experts tend to use solution synthesis strategies to plan and execute a search of the problem space and enable the solver to infer a solution. In realistic transmission genetics problems the solution strategy that is used by experts is hypothesis testing. Experts formulate two types of hypotheses -- general hypotheses about the inheritance patterns and modifiers and specific hypotheses about the distribution of variations to offspring for each cross. Because new data is continuously produced, there is an interaction between the problem data, the specific hypotheses and the general hypotheses. One expert begins:

> "I've got four classes each of males and females so there is no reason not to think it is simple so I'll cross the dw's (dumpy wing, white eye) with the sc's (shiny wing, cinnabar eye) and all the offspring are dw (dumpy wing, white eye), so if d (dumpy wing) and w (white eye) are dominant, the offspring are all heterozygotes..."

In the example, the initial population data presents an organism with two variations for each of two traits. The redescription allows the expert to retrieve the knowledge to formulate an initial, tentative general hypothesis of simple dominance. The expert then chooses to cross parents with unlike variations, using the specific hypothesis that if the genotype of one parent is homozygous dominant and the genotype of the other parent is homozygous recessive, the offspring will be heterozygous having a dominant phenotype, to predict the distribution of variations among the offspring. This cross is then performed, and the results agree with the prediction. The newly generated data supports the specific hypothesis and the specific hypothesis helps the solver infer the general hypothesis. This interaction between data, specific hypotheses, and general hypotheses continues throughout the synthesis of the problem solution.

Also, in the solution synthesis, for each inheritance pattern and modifier, there is a cross or class of crosses that, once performed and explained, assures the solver that the solution is justifiable. This cross is being termed the definitive cross. In simple dominance and codominance this definitive cross is the F(2) cross; in multiple alleles the class of crosses used to justify the solution includes two F(2) crosses. An F(2) cross is between two parents that are known to be heterozygotes with the distribution of variations to the offspring in a 3:1 (dominant:recessive) ratio. In the example begun earlier in this paragraph the expert continues solving the problem by using the offspring from the first cross, assuming they are heterozygotes, as parents in the second cross. This is an F(2) cross for both traits. The definitive cross in all classes of problems except sex linkage requires the identification of heterozygous individuals. In this problem the expert has constructed heterozygous individuals by crossing parents with unlike phenotypes.

Once the inheritance pattern has been inferred, the expert continues to do crosses to decide if a modifier is operating on the population. Either because of indicators in the problem data and/or to assure themselves the solution is complete, experts usually consider both sex linkage and autosomal linkage modifiers. In testing for modifiers, the interaction between the problem data, the specific hypotheses, and the general hypotheses continues. There is also a definitive cross to justify each modifier. In sex linkage the definitive cross is between a dominant male and a recessive female, producing recessive male and dominant female offspring. In the two-trait autosomal linkage problems, the definitive cross is between a parent that is heterozygous for both traits and another that is homozygous recessive for both traits. The indication that the traits are not independent is that the ratio of the distribution of the variations to the offspring is not the expected 1:1:1:1 ratio.

By formulating two types of hypotheses, and by generating additional data that are either explained by, or predicted from, an hypothesis, experts are able to infer justifiable solutions to genetics problems.

Solution Assessment Experts tend to use solution assessment strategies to assure the solver that the solution is as complete and accurate as possible. While determining the presence of a modifier in the problem, the experts are assuring themselves that the solution to the problem is complete.

Experts confirm that a solution is accurate by collecting additional evidence beyond the definitive cross. Although the Chi square test can be used to determine if the observed distribution of variations to offspring agrees with the expected distribution, experts seldom use the Chi square test. Rather, they compare the ratios of the distribution of the variations by intuition, without the formal statistical test. Experts also increase their confidence in the accuracy of the inheritance pattern and modifier hypotheses by doing additional crosses that are explained by or predicted from the general and specific hypotheses. Whenever possible, experts use more than one method of confirmation. One example of confirmation is,"I think now I'll do its reciprocal." Another expert says, "...this is basically the 9:3:3:1 - 20:9:5:2, which is very, very, very close. So I'm sure I know what is going on already. Might as well confirm it by a test cross." A third example of confirmation is the expert who says, "I think I'll just repeat that cross a few times to jack up the numbers before I pull out my calculator...Oh, the ratio is getting closer all the time." The description of the strategic knowledge of experts used to solve introductory level realistic transmission genetics problems is summarized in Table 7.

Table 7

Summary of the Characteristics of Strategic Knowledge

---

1. Data Redescription

   -- Consists of

      -- number and name of variations
      -- number and name of traits
      -- number of classes of phenotypes
      -- missing classes of phenotypes
      -- unequal distribution of individuals to classes of phenotypes
      -- initially occurs prior to formulation of a general hypothesis

---

2. Solution Synthesis

   -- Consists of hypothesis testing

      -- general hypotheses about inheritance patterns and modifiers
      -- specific hypotheses about crosses

   -- Occurs by

      -- using hypotheses to explain data generated by crosses
      -- predicting new data by crosses from hypotheses

   -- Requires

      -- interaction of data, specific hypotheses and general hypotheses
      -- performing a definitive cross using heterozygotes

---

3. Solution Assessment

   -- Consists of confirmation -- specific to selected inheritance patterns and modifiers

   -- Occurs by collecting additional evidence

      -- through Chi square and other informal mathematical tests
      -- by doing additional crosses

   -- Includes more than one form of confirmation if possible

---

Table 8 summarizes the genetics feature of each category of strategic knowledge used by the experts to infer the solution for each class of problems.

The description of the performance of experts solving realistic computer-generated transmission genetics problems can also be summarized as a flowchart (Figure 3). In this flowchart there are many paths and feedback loops, but the three categories of strategic knowledge used in solving genetics problems--data redescription, hypothesis testing and confirmation--regularly recur. From the flowchart it is also evident that the opportunity to produce problem data is essential for the solution of these realistic problems.

Implications

From the description of experts' knowledge, one implication can be made about the utility of using Reif's model as a starting point for the study of problem solving in science. The categories of strategic knowledge identified by Reif to describe problem solving in physics -- data redescription, solution synthesis and solution assessment -- have been used to describe problem solving in transmission genetics. The details within each category are different for genetics problems and physics problems, but this is expected since the disciplines are different, and the realistic problems studied in genetics are not like the textbook problems studied in physics either in structure and form. Among the differences are:

1) that in the physics problems the data is limited to what is given in the problem statement while in the genetics problems continuous data production is possible;

Table 8

Summary of Details of Strategic Knowledge

	REDESCRIPTION	SOLUTION SYNTHESIS	SOLUTION ASSESSMENT
	CHARACTERISTICS	DEFINITIVE CROSS	CONFIRMATION
Simple Dominant	2 variations/ trait	F(2)	Chi square linkage
Co-Dominant	3 variations/ trait	F(2)	Chi square linkage
Multiple Alleles	3-6 variations /trait	Series of crosses with an F(2)	Match all phenotypes to a genotype
Sex Linkage	Missing class of phenotype of one sex	Dominant m X recessive f	None
Autosomal Linkage	Missing or low frequency class of phenotypes	Linkage	Repeat cross with different individuals

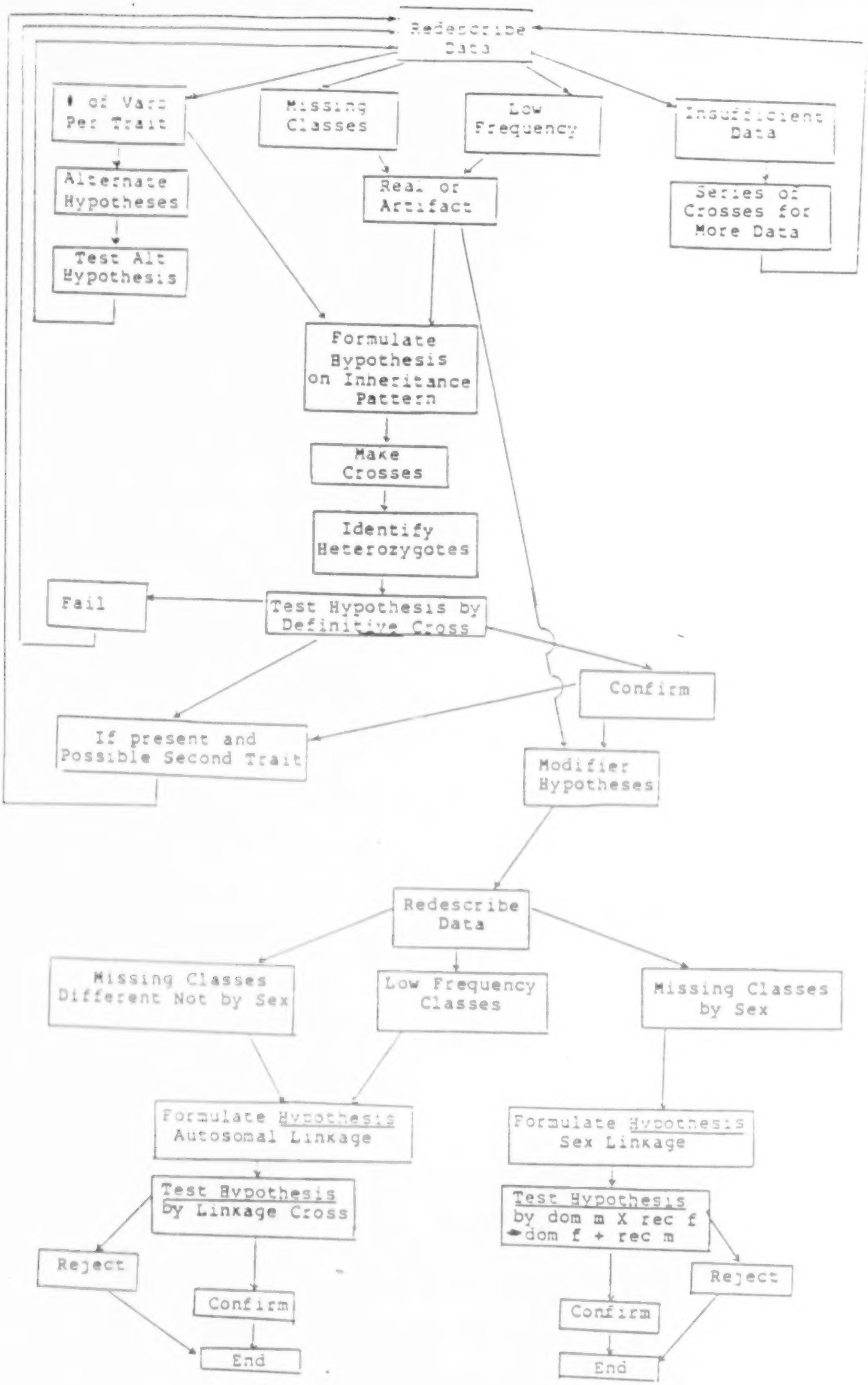

Redescribe Data
# of Vars Per Trait
Missing Classes
Low Frequency
Insufficient Data
Alternate Hypotheses
Real or Artifact
Series of Crosses for More Data
Test Alt Hypothesis
Formulate Hypothesis on Inheritance Pattern
Make Crosses
Identify Heterozygotes
Fail
Test Hypothesis by Definitive Cross
Confirm
If present and Possible Second Trait
Modifier Hypotheses
Redescribe Data
Missing Classes Different Not by Sex
Low Frequency Classes
Missing Classes by Sex
Formulate Hypothesis Autosomal Linkage
Formulate Hypothesis Sex Linkage
Test Hypothesis by Linkage Cross
Test Hypothesis by dom m X rec f → dom f + rec m
Reject
Confirm
End
Reject
Confirm
End

Figure 3. Flowchart of Solution Path used by Experts to Solve Realistic Transmission Genetics Problems

2) that in the physics problems the solution requires a mathematical formula while no mathematical formula exists for the solution of the genetics problems; and

3) that in the physics problems the solution has a numerical value while in the genetics problems the solution is a confirmed hypothesis.

In light of the dissimilarity of physics and genetics problems, the fact that the same categories of strategic knowledge can be used to describe problem-solving performance in both disciplines supports the utility of the model.

A second implication, which may be important both to the study of problem solving and to the design of instruction in problem solving in science, is about the content knowledge of expert problem solvers in genetics. Although content knowledge is not the emphasis of this study, it is evident that experts have a large store of highly organized, easily retrievable information available for problem solving. The use of strategic knowledge could not be described without reference to the content knowledge -- for example, of inheritance patterns and modifiers, of specific crosses, of traits and variations, of dominant and recessive variations, of phenotypes and genotypes, of homozygotes and heterozygotes. It is also evident that this content knowledge has associated with information of when and how to use the strategic knowledge. For example, the experts know that an F(2) cross yields data useful in testing the simple dominant inheritance pattern hypothesis, and that this cross requires heterozygous individuals. In the study of problem solving, further research is needed to analyze and explicate the content knowledge required for successful problem solving in genetics. Likewise, instruction designed to teach problem-solving strategies cannot be independent of instruction in the content of the discipline.

Conclusion.

The advent of realistic, computer-generated problems has created opportunities for students to achieve important learning outcomes in science. As models for understanding and teaching problem solving develop and as technology makes the computer a powerful and available instructional tool, science educators need to continue to design instruction to provide students with improved learning experiences in problem solving. One step toward achieving the goal of improved instruction and learning in problem solving is to describe the performance of successful problem solvers.

REFERENCES

Bhaskar, R., & Simon, H. A., (1977). Problem solving in semantically rich domains: An example from engineering thermodynamics. Cognitive Science, 1, 192-215.

Chi, M. T. H., Feltovich, P. J., & Glaser, R., (1981). Categorization and representation of physics problems by experts and novices. Cognitive Science, 5 (2), 121-152.

Elstein, A., Shulman, L., & Sprafka, S. (1978). Medical problem solving: An analysis of clinical reasoning. Cambridge: Harvard University Press.

Jungck, J., & Calley, J. (1986). Genetics: Strategic simulations in Mendelian genetics - User's manual, COMPress, Wentworth, N. H.

Larkin, J. H., McDermott, J., Simon, D. P., & Simon, H. A. (1980). Expert and novice performance in solving physics problems. Science, 208, 1335-1342.

Larkin, J. H., & Rainhard, B. (1984). A research methodology for studying how people think. Journal of Research in Science Teaching, 21 (3), 235-254.

Larkin, J. H., & Reif, F. (1979). Understanding and teaching problem solving in physics. European Journal of Science Education, 1 (2), 191-203.

National Academy of Science & National Academy of Engineering. (1982). Science and Mathematics in the Schools: A Report of a Convocation, Washington, DC: National Academy Press.

Reif, F. (1983a, June). Understanding and teaching problem solving in physics. *Lectures at the International Summer School on Physics Education*, LaLonde Les Maures, France.

Reif, F. (1983b). How can chemists teach problem solving? *Journal of Chemical Education*, *60* (11), 948-953.

Smith, M. (1983). *A comparative analysis of the performance of experts and novices while solving selected classical genetics problems*. Paper presented at the meeting of the National Association of Research in Science teaching, Dallas, TX.

Smith, M., & Good, R. (1984a, April) *A proposed developmental sequence for problem-solving ability in classical genetics: The trial and error to deductive logic continuum*. Paper presented at the meeting on the National Association of Research in Science Teaching, New Orleans, LA.

Smith, M., & Good, R. (1984b). Problem solving and classical genetics: successful vs. unsuccessful performance. *Journal of Research in Science Teaching*, *21* (9), 895-912.

Stewart, J. H., (in press). Potential learning outcomes from problem-solving instruction: Examples from genetics. *Science Education*.

APPENDIX C

MENDEL Research Report #2

MENDEL Research Report #2

High School Students' Problem-Solving Performance on Realistic Genetics Problems

Susie Johnston Slack
James Stewart
March 1988

Work on Project MENDEL has been supported by the Dean of the School of Education - University of Wisconsin-Madison; the IBM Corporation; and the US Department of Education's FIPSE program (Grant #G008301471). Specific support for the research reported here came from a National Science Foundation Grant (MDR8470277).

THE AUTHOR'S ADDRESSES ARE:

Susie Johnston Slack
James Stewart
226 TEB
University of Wisconsin-Madison
225 N. Mills Street
Madison, Wisconsin 53705

High School Students' Problem-Solving Performance on Realistic Genetics Problems

Abstract

Problem solving is recognized as a valuable educational experience in science. Thus genetics, essentially a problem solving science included in almost all high school biology courses offers a fruitful area for studying student problem solving performance. The research reported in this paper describes the performance of 30 high school students solving 119 problems generated by the computer program GENETICS CONSTRUCTION KIT (Jungck and Calley, 1985). Solving GCK problems requires students to plan experiments, generate and interpret data, and reason from causes (phenotypic data) to effects (genotypic data). Research data consisted of transcribed audio-tapes of students thinking aloud as they solved problems and computer print-outs of initial data and sequence of crosses. Transcripts were analyzed for common actions and comments made during the problem solving process in terms of: initial data redescription and interpretation, hypothesis generation, cross data redescription and interpretation, solution, and solution confirmation. This study was done in an effort to add to the understanding of student problem solving strategies and to develop a model of student performance. A model, that when combined with a model of expert performance may serve as a basis for improving genetics instruction.

High School Students' Problem-Solving Performance on Realistic Genetics Problems

Introduction

> There is wide acceptance that one important goal of [science] education is to give students experience in solving problems. If this is accepted, then it is necessary to deal with problem solving, and not just with solutions. (Moore, 1985)

This statement underscores the importance of the current emphasis on problem solving as a valuable educational goal in science. There has been a great deal of research done on problem solving with the expectation that it will lead to improved instruction. Much of this research has been on physics content, where it has been shown that experts and novices structure their knowledge very differently (Larkin, McDermott, Simon, & Simon, 1980; Chi, Feltovich, & Glaser, 1981), and that the structuring of that knowledge influences the manner in which problems are solved. Reif (1983) has pointed out that to improve instruction, researchers need to develop models of desired performance (derived from descriptions of expert performance and theoretical analyses of disciplines), of novice performance, of learning, and of teaching.

Problem-solving research has also been done in genetics. For example, Smith and Good (1984) identified 32 problem-solving tendencies used by successful, but not unsuccessful, solvers. They observed that strategies of successful solvers included seeking a solution rather than an answer, checking for consistent logic, working forward, checking for one variable (trait) at a time, and looking for evidence that would invalidate previous assumptions. In related research, Hackling (1984; 1986) described the performance of experts and novices solving pedigree problems. He concluded that while experts did not differ from novices in the number of correct answers obtained, they were able to justify their solutions in terms of underlying genetics concepts at a level which novices could not achieve. Experts identified more critical cues about genetic mechanisms, generated and tested hypotheses, considered alternate hypotheses, and recognized the need to modify strategies based on problem conditions. These findings are similar to those reported by Smith (1986).

At the University of Wisconsin-Madison we have studied high school students solving textbook genetics problems (Stewart, 1983) and the mental models of meiosis that underlie their problem-solving performance (Stewart & Dale, 1988). We have also been using realistic problems generated by the microcomputer program GENETICS CONSTRUCTION KIT (GCK), developed by Jungck and Calley (1985), as a vehicle to study problem solving in genetics. For example, Albright (1987) has studied the problem-solving performance of university genetics students and Collins (1986; 1987) has studied PhD geneticists as they solved GCK problems. By combining the results of this research with her own analysis of the structure of transmission genetics, Collins has developed a model of desired performance for transmission genetics.

We have focused our research on GCK problems because they require students to both generate and interpret data, and are thus more likely than textbook problems to lead to important learning outcomes (Stewart, 1988). However, we agree with Kinnear (1983a; 1983b) that, while computer simulation environments make it possible to offer students experiences to develop their understanding of genetics, they do not in and of themselves improve problem-solving performance. Research such as that reported in this paper can add to the understanding of novices and lead to the development of a model of student performance. Such a model, when combined with results of related research on expert performance, will lead to improved genetics instruction.

The Study

Thirty students, from five high schools, took part in this study. These students, in grades 9-12, had completed three to four weeks of genetics instruction in introductory level biology courses and were selected by their teachers because they represented a range of ability and grade levels and because they were judged likely to think aloud as they solved problems.

The Problems

The problems used in this research were produced by a version of GCK that kept records of each student's interactions with the program. Problems were selected to be consistent with the genetics that the students had been taught. The 30 students solved a total of 119 problems. Details of the number of problems of each type solved by students is shown in Table 1.

Table 1 Here

Using GCK, researchers can create problem classes from which instances are generated for students to solve. It is possible to construct monohybrid through tetrahybrid problems with combinations of the inheritance patterns simple dominance, codominance, multiple alleles, and gene interaction, and the modifiers sex linkage, autosomal linkage, lethality, penetrance, and pleiotropy. Each problem begins with a population of field-collected organisms, with the sex and phenotype of each individual identified. Once the field collection is displayed, students can produce offspring data by selecting individuals to be the parents for crosses. Generations of offspring can be produced until a student is ready to explain the phenotype data in terms of inheritance patterns and modifiers. Therefore, to obtain a solution, a student must plan experiments, make crosses, and interpret data. Solving these problems requires students to reason from effects (phenotype data) to causes (underlying genetics mechanisms). Typical textbook problems require reasoning in the reverse direction.

Data Gathering

During problem-solving sessions with individual students, two types of research data were gathered -- audio taped think aloud protocols (which were subsequently transcribed) and printouts of information on the initial population and the sequence of crosses performed. In the first session, each student was given instruction on how GCK worked, then solved one monohybrid simple dominance practice problem. This was done to help them feel comfortable with thinking aloud, with the researcher, and with GCK. As part of the practice problem students were introduced to the concept of a field collection and it was noted that the initial sample of field data displayed on the screen was analogous to vials of randomly collected flies from a wild population. Throughout the session researchers reminded students about the concept of a field collection when it seemed necessary.

Following the practice problem, each student solved three or four more problems during the remainder of the first and during a second problem-solving session. Researchers intervened in the solving process only to encourage students to think aloud.

Data Analysis

The data analysis was done in two stages. First, the transcripts of twelve students were reviewed and a list was made of their actions and comments within the data analysis categories of data redescription, hypothesis generation, crosses performed (phenotypes and individuals used as parents), solution given, and the method of confirmation used. These categories were those established by Collins (1986). Second, the remaining 18 transcripts and the original 12 were reviewed noting the occurrence of these common actions and comments. Transcripts were then matched with the appropriate printouts of a student's sequence of crosses. Each transcript was divided into all comments made between one cross and the next one, beginning with those made between the presentation of the initial population and the first cross.

All monohybrid simple dominance problems were analyzed first. Dihybrid and codominant problems were then analyzed in the same way to allow for comparisons within and across problem types. The analysis imposed a consistent framework on the solution-seeking process so that a general description of student problem-solving performance could be made. Parts of a sample transcript, analysis, and corresponding problem data are included in Figure 1. This analysis procedure made it possible to describe students' problem-solving performance in terms of:

1. initial data interpretation,
2. factors in the initial data that influenced the first cross,
3. when, and on what basis, hypotheses were generated,
4. the consistency of hypothesis usage as a basis for generating, explaining and predicting data,
5. planning processes,
6. the means (qualitative or quantitative) that students used to interpret data,
7. the warranted, unwarranted, and missed warranted inferences, and
8. the nature of the justifications and solution confirmation procedures.

Figure 1 Here

The results reported represent a composite of the actions and responses of 30 students across all problems. These actions and responses occurred in both correctly solved and incorrectly solved problems. This section is organized around the major activities involved in the solving of the GCK problems: initial data redescription, interpretation, and hypothesis generation; initial cross and cross strategies; cross data redescription, cross interpretation, hypothesis generation; solutions; and, solution confirmation.

Initial Data Redescription, Interpretation, and Hypothesis Generation

Student redescription of initial data included reading data from the screen, repeating the names and numbers of individuals, and noting the least or most frequent phenotype. For example, a student might say "There are 16 red females and 44 yellow females and 36 red males and 10 yellow males. There are more yellow ones altogether". Students commented on unequal numbers of males and females as possible clues to sex-linkage. They also counted and added numbers of individuals of a phenotype in order to compare numbers in the data with known ratios, for example that 3:1 is indicative of simple dominance. If the ratios were close, students would develop hypotheses based on the initial population. However, when numbers did not reflect ratios known to the students, they generally ignored them. Typically, students did not summarize data in terms of numbers of traits and variations, using such qualitative redescription to make hypotheses about possible inheritance patterns.

Two unwarranted inferences on which tentative hypotheses were based were made during initial data interpretation. First, students assumed that the most frequent phenotype was dominant ("There are more yellow ones altogether so yellow is dominant"). Second, students interpreted slightly unequal numbers of males and females in a population as an indication of sex-linkage, or conversely the existence of both male and female phenotypes as indicating no sex-linkage. For example, a student might say "There are fewer orange males than yellow males, so orange might be sex-linked"; or "Well, there are males and females of both variations, so this problem doesn't have sex-linkage in it."

Students generated hypotheses from the initial data on a phenotypic level and expressed them in terms of a trait [sic] being dominant or recessive. Even though the concept of a field collection had been carefully explained in all introductory sessions and throughout both problem-solving sessions, some students had difficulty understanding that the initial data was only a random sample of a larger population. Consequently students often tried to solve problems from the initial data rather than using the data as a starting point for solving the problem.

Initial Crosses and Cross Strategies

When a student stated a hypothesis based on the initial data, the first cross was not necessarily planned from the stated hypothesis. Further, students did not systematically perform back crosses, test crosses, or reciprocal crosses to test hypotheses. Rather, they normally used individuals from the initial population for their crosses.

As a result, they did not normally create family lines, which would have been the case if offspring generated from previous crosses had been used as parents.

There were three common approaches to making crosses:

1. The initial cross was done to "see what happens" or "just to do something". Subsequent crosses appeared to be done to make crosses of all possible combinations of parental phenotypes.
2. The initial cross was between parents of the same phenotype, in order to find two heterozygous individuals -- to cause the hidden variation to "show up". In this approach, the initial hypothesis, although often based on the unwarranted inference that the most frequent phenotype was necessarily dominant, did influence the first cross. A student would comment "There were more reds and I think red is dominant. I'll cross two reds and if any yellows show up I'll know I'm right". Students often made several crosses with parents of the same phenotypes to discover heterozygous individuals. If heterozygous individuals were discovered, students would cross two individuals with the phenotype believed to be recessive. This approach was useful and provided information that the students could interpret. When students had repeatedly crossed individuals with the same phenotype and failed to discover the hidden variation, two basic strategies were used: 1) crossing of individuals with unlike phenotypes in an attempt to prove dominance, expecting the offspring

produced to be of one phenotype or 2) crossing parents of the same phenotype (the one not used in their original cross) several times hoping that the hidden variations would be revealed.

3. A third initial cross involved using parents with unlike phenotypes. Students expected the offspring produced from this cross would be of one phenotype, thus indicating the dominant variation. They would comment "If I cross a red and a yellow and get all red offspring, I'll know red is dominant". If these results were obtained, their second cross would be between like phenotypes that had been "hidden" or were recessive. A student would comment "Now I know red is dominant, so if I cross two yellows I should get all yellows". When these results were obtained, students would repeat a cross of unlike phenotypes with different individuals to further substantiate that a trait was dominant. When the results of the second unlike cross duplicated the results of the first unlike cross students were confident of their conclusions. However, when a second cross with individuals of unlike phenotypes produced conflicting results (that is, the offspring were of two phenotypes), students became confused: "How can parents that are the same produce different offspring?"

Cross Data Redescription, Interpretation, and Hypothesis Generation

Data redescription following crosses consisted of students reading the names of parents crossed and the offspring produced, noting the least and most frequent phenotypes, and counting total numbers of each phenotype produced. It was not uncommon for students to perform two or three crosses before redescribing the data. Redescriptions that occurred after several crosses were done to summarize data into some pattern on a phenotypic level and to generate hypotheses. This type of redescription and hypothesis generation often included such statements as "In cross #1 the parents were hooked and lyra and all the offspring were lyra. That same thing happened in cross #3. I think hooked is recessive." Summarizing cross data redescription also was done when a cross produced data that seemed to conflict with a previous cross. For example: "In cross #1 I crossed a hooked and a lyra and all the offspring were hooked. In cross #3 the same thing happened. I thought hooked was dominant to lyra. But now in cross #5 I crossed the same kinds of parents and got offspring of both." In some cases, students chose to ignore data that conflicted with what they could explain or interpret saying, "That last cross must have been a mutation or mistake, because the parents were the same." Cross data redescription also involved counting the number of individuals by phenotype and interpreting the numbers in terms of known or expected ratios. As was the case with initial data redescription, students ignored numbers that did not exactly fit expected ratios.

Students missed inferences or made unwarranted inferences during cross data interpretation, with the most common class of missed inferences being failure to recognize that cross results provided new and valuable information. Missed inferences:

1. in simple dominance problems, not inferring dominance when a cross between parents of unlike phenotypes produced offspring of one phenotype.
2. in simple dominance problems, not inferring dominance when a cross between parents of like phenotypes produced offspring of two variations.
3. in codominance problems, not inferring the codominant phenotype (heterozygote) when a cross between parents of different phenotypes produced offspring of three variations.

Examples of unwarranted inferences (those not justifiable given the data) common to all problem types in addition to those noted in initial data interpretation included:

1. crossing individuals with like phenotypes, producing offspring of all one phenotype, and inferring dominance when the variation could just as well have been recessive.
2. crossing individuals of the most frequent phenotype and inferring that the variation produced in greatest numbers was the dominant variation.

After crosses, students used their hypotheses to explain data. As a result of missed and unwarranted inferences made during data interpretation, students often changed their hypotheses from cross to cross, acting as if each cross was a separate problem. As with initial hypotheses, these were made on a phenotypic level and based on a small amount of data produced from one cross independent of other data. For example, a student might cross two red parents and assume that because all offspring are red that red is dominant. Then, in the next cross, two yellow parents might be crossed; because all the offspring are yellow, the hypothesis would be changed to say that yellow is dominant.

Students did not clearly distinguish between inheritance patterns and modifiers in their hypotheses. This was especially evident in students' considerations of sex-linkage. As in initial data interpretation, students considered or discounted sex-linkage as an inheritance pattern, not a modifier, by comparing the number of males with females produced in a cross. Sex-linkage was not considered when numbers of males and females were equally produced in cross data.

Solutions

Students' solutions, like their hypotheses, emphasized phenotypic interpretations with little or no genotypic explanation. Solutions also included inaccurate use of allelic symbols and a failure to distinguish between inheritance patterns and modifiers. While students solved (got the correct answers for) monohybrid simple dominance problems, solutions to dihybrid and codominance problems proved more difficult. Students decided a problem with simple dominance inheritance was solved correctly when they could identify dominant and recessive phenotypes. Justification for solutions, some based on the unwarranted inferences previously described, included: crossing parents of the same variation, producing offspring of one variation and concluding that variation was recessive; crossing parents of two variations, producing offspring of one variation, and concluding the variation of offspring produced was dominant; crossing parents of two variations and concluding the variation produced in the offspring in greatest numbers was dominant; and crossing parents of two variations and concluding a variation was dominant because numbers of offspring produced fit a 3:1 ratio. Solutions for dihybrid problems with simple dominance inheritance were solved using the same justifications given for monohybrid problems. Students solved dihybrid problems one trait at a time.

When they were unable to explain data in codominance problems, students invented new inheritance patterns with new genotype-to-phenotype relationships, such as co-recessive and double dominance, to support the patterns. In the co-recessive explanation, three letter symbols were used for the three variations, as if each phenotype were the expression of separate alleles. For example, a student's explanation of a cross between two individuals with the same phenotype that resulted in offspring of three phenotypes was that two variations were recessive to the third. Double dominance was explained using three letters as if two variations were dominant to the third. Thus, when two dominant phenotypes were crossed, the third variation would be produced as the "hidden recessive" trait, an explanation that had been useful in simple dominance problems.

Solution Confirmation

Students did not actively confirm solutions by generating additional data or by using any statistical tests, such as a Chi Square.

They did not recognize the necessity of confirming the accuracy and completeness of a solution. The only method used to confirm a cross was to repeat the last cross performed, using the same individuals as parents. Occasionally, students did draw a Punnett square diagram on scratch paper to confirm a solution. However, diagrams only confirmed one cross. Some students also redescribed several crosses, looking back over accumulated data and selecting crosses that were consistent with their hypotheses. They used ratios to confirm their solutions concerning which variations were dominant or recessive. As in data interpretation, if the ratios were not exact, the students would change their solutions.

Discussion

Three trends in general problem-solving procedures were evident in our results. These trends were: (1) an unplanned approach, characterized by a lack of specific hypothesis generation and testing; (2) working backward -- explaining cross data rather than predicting it; and (3) emphasis on a quantitative level of counting numbers of individuals and using ratios in individual crosses.

In addition to these general problem-solving trends, there was evidence that most students lacked three important genetics-specific ways of thinking about problem solving: genotypic thinking; generational thinking; and ability to distinguish between an inheritance pattern and a modifier. Each of these trends was substantiated by actions and responses throughout the problem-solving process.

Indications of a lack of genotypic thinking included:

1. being unable to explain why two crosses of parents of the same phenotype could produce offspring with different numbers of variations. In other words, students did not understand that a single phenotype could map to two genotypes.
2. changing hypotheses from cross to cross. Notably, when a cross between individuals of like phenotypes (homozygous dominants) was followed by a cross between like phenotypes (homozygous recessives), students could not decide which variation was dominant.
3. inventing phenotype-to-genotype relationships with symbolic representations in explanations of codominance.

Evidence that students did not think generationally included:

1. employing a cross strategy using parents, primarily from the initial population, to search for a heterozygote to produce a "hidden trait" rather than creating a heterozygote or lines of familial data so that the exact heritage (genotype) of parents could be established.
2. considering each cross as a separate problem, as though the dominance and recessive variations could change from cross to cross within a problem.

Evidences of students' lack of distinguishing between inheritance patterns and modifiers included:

1. failing to qualitatively redescribe data in terms of the numbers of traits and variations.
2. attending to information in data that was potentially misleading, such as unequal number of males and females of a pheno-

type, or equal representations of males and females in a phenotype as indicating sex-linkage.

Educational Implications

The use of computer simulations to promote the acquisition of problem-solving skills in genetics and other sciences is likely to increase in the coming years. However, it is our belief that simply providing an environment in which students are presented realistic genetics problems is not sufficient to elicit good problem-solving skills. Research is needed to develop models of problem-solving performance that can be used to develop instruction. Such instruction will help students develop explicit connections between conceptual knowledge and problem solving (including both content-independent and discipline-specific problem-solving strategies). From the research reported here, we recommend that instruction in genetics be designed so that:

1. Genetics concepts and principles are presented in such a way that explicit relationships between concepts (i.e. chromosomes, genes, alleles, traits, and variations) are obvious.
2. An important feature of teaching about inheritance patterns is the use of the number of variations per trait as a clue to possible inheritance patterns.
3. Qualitative redescription is taught in terms of clues or patterns that lead directly to a tentative hypothesis about an inheritance pattern and solution to a problem.

4. Hypothesis generating and testing is taught as the strategy used to solve a problem, since an hypothesis provides direction as to what crosses to do and how to interpret the results of crosses.
5. Students are taught to make hypotheses and to select crosses from previous generations, not from initial field data alone, in order to create familial lines of data about which genotypes may be established.
6. Students are taught to understand the relationship between genotype and phenotype as a basis for understanding inheritance patterns.
7. Students are taught the importance of expressing a solution in terms of an inheritance pattern, and of checking or verifying the solution for accuracy and completeness.

These suggestions for instruction in genetics reinforce problem-solving strategies that can not only be used in genetics, but in other disciplines as well. They include predicting data, redescribing a problem qualitatively, generating and testing hypotheses, considering alternative hypotheses, and checking results. In order for problem solving to be a valuable educational experience, it is necessary to teach not only conceptual knowledge, but the relationship of conceptual knowledge to problem solving. It is likely that students will improve their problem-solving skills and their conceptual knowledge of genetics, as well as gain a better understanding of the nature of science if, during instruction, the relationships of problem solving to genetics concepts and problem solving as decision-making are stressed.

Albright, W. C. (1987) A Description of the Performance of University Students Solving Realistic Genetics Problems. Masters Thesis, University of Wisconsin-Madison.

Chi, M.T.H., Feltovich, P.J. & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. Cognitive Science, 5, 121-152.

Collins, A. (1986) Strategic knowledge required for desired performance in solving transmission genetics problems. Ph.D. Dissertation, University of Wisconsin-Madison.

Collins, A. (1987). A description of the strategic knowledge of experts solving transmissions genetics problems. Manuscript in Review.

Hackling, M. W. (1984). Expert and novice performance in solving genetics pedigree problems. Paper presented at the Annual Conference of the Australian Association for Research in Education, Perth, Western Australia, November, 1984.

Hackling, M. W. (1986). Hypothesis testing strategies used in the solution of genetic pedigrees. Paper presented at the Annual Meeting of the American Educational Research Association. San Francisco, CA. April 1986.

Jungck, J. R. and Calley, J. N. (1985) Strategic simulations and post-Socratic pedagogy through experimental inquiry. American Biology Teacher, 47, 11-15.

Kinnear, J. F. (1983a). Identification of misconceptions in genetics and the use of computer simulations in their correction. In H. Helms and J. D. Novak (Eds) Proceedings of the International Seminar on Misconceptions in Science and Mathematics, 84-92, Ithaca, NY Cornell University.

Kinnear, J. F. (1983b). Using computer simulations to enhance problem-solving skills and concept development in biology students. Proceedings of the Conference on Computer-Aided Learning in Tertiary Education University of and, Queensland 4067, Australia 189-201.

Larkin, J., McDermott, J., Simon, D.P., & Simon, H. (1980). Expert and novice performance in solving physics problems. Science, 208, 1335-1342.

Moore, J. (1985). Science as a way of knowing. Report of the committee on education of the American Society of Zoologists. Baltimore, MD.

Reif, F., (1983). Understanding and teaching problem solving in physics. Lectures at the International Summer School on Physics Education, La Londe Les Maures, France.

Smith, M., (1986). Problem solving in classical genetics: successful and unsuccessful pedigree analysis. Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco, CA. April 1986.

Smith, M. and Good, R. (1984). Problem solving and classical genetics, successful versus unsuccessful performance. Journal of Research in Science Teaching, 2, 895-912.

Stewart, J. (1983) Student problem solving in high school genetics. Science Education, 67, 523-540.

Stewart, J. (1988) Potential learning outcomes from solving genetics problems: A topology of Problems. Science Education. (In press).

Stewart, J. and Dale, M. (1988). High school students understanding and chromosome/gene behavior durig meiosis. Manuscript in review.

Acknowledgments

The research reported on in this paper was made possible by a grant (# MDR 8470277). We would also like to thank the following colleagues for comments made on various drafts of this paper: William Albright, Angelo Collins, Peter Hewson, Norman Thomson, Steven Slack, and Judith Van Kirk.

Table 1

Number of Problems Solved by Problem Type

School	# students	Grade	MSD	MCD	DSD	DCD
1.	12	10-12	25	18	17	0
2.	3	9	3	2	2	3
3.	6	9	6	5	5	2
4.	6	11-12	12	5	3	4
5.	3	9	2	0	0	5
Total	30		48	30	27	14

KEY

MSD: Monohybrid with Simple Dominance

MCD: Monohybrid with Codominance

DSD: Dihybrid with Simple Dominance only

DCD: Dihybrid with Codominance and Simple Dominance

Figure 1

Sample Transcript and Cross Data

OK, we're dealing with three traits. [Redescription, Bold-Researchers analysis comment]	Field Collection Vial #1 20 F Dumpy 13 M Dumpy 6 F Inverted 8 M Inverted 8 F Expanded 9 M Expanded
So let's see...I'll take a female dumpy individual from Vial 1 and a male dumpy individual from Vial 1. [Cross #1]	Cross #1 Vial #1 Dumpy F Individual #1 X Vial #1 Dumpy M Individual #1
OK. We got offspring of all three kinds. [Cross data redescription]	Contents of Vial #2 7 F Expanded 5 M Expanded 13 F Dumpy 16 M Dumpy 8 F Inverted 4 M Inverted
So now, I'll take some more offspring, or, parents from vial 1 this time, I can't remember what the I stood for. Oh well, we'll take some of them anyway. So I'll take some of the I's female and a male. [Cross #2]	Cross #2 Vial #1 Inverted F Individual #1 X Vial #1 Inverted M Individual #1

OK, when we cross the two inverted, we got all inverted offspring. [Cross data redescription]

Contents of Vial #3

27 F Inverted

26 M Inverted

OK, for this problem I'm trying to figure out what's dominant and what's recessive. Let's see we got all inverted. So now I'll cross some of the E's to find out what they are, from Vial 1 again. [Cross #3]

Cross #3

Vial #1 Expanded F Individual #1

X

Vial #1 Expanded M Individual #1

OK, When we crossed E's we got all E's for Offspring [Cross data redescription]

Contents of Vial #4

34 F Expanded 34 M Expanded

So let's see, I'm gonna try dumpy parents from Vial 2 this time. To see if the results are any different from the parents in Vial one. [cross four: repeat phenotypes]

Cross #4

Vial #2 Dumpy F Individual #1

X

Vial #2 Dumpy M Individual #1

Contents of Vial #5

14 F Inverted	10 M Inverted
12 F Expanded	4 M Expanded
8 F Dumpy	13 M Dumpy

APPENDIX D

MENDEL Research Report #4

MENDEL Research Report #4

High School Students' Understanding of Chromosome/Gene Behavior During Meiosis

Jim Stewart
Michael Dale
March 1988

Work on Project MENDEL has been supported by the Dean of the School of Education - University of Wisconsin-Madison; the IBM Corporation; and the US Department of Education's FIPSE program (Grant # G008301471). Specific support for the research reported here came from a National Science Foundation Grant (SED 8022912).

The authors' addresses are:

Jim Stewart
Michael Dale
226 TEB
University of Wisconsin-Madison
225 N. Mills Street
Madison, Wisconsin 53706

High School Students' Understanding of Chromosome/Gene Behavior During Meiosis

Introduction

During the past decade, two research emphases have emerged in science education -- work on alternate conceptions (for example, Driver & Easley, 1978; and Erickson, 1979; Hackling & Treagust, 1982, 1984; Kargbo, Hobbs & Erickson, 1980); and work on problem solving (for example, Chi, Feltovich & Glaser, 1981, Larkin, 1983; Reif, 1983; Smith & Good, 1984; and Stewart, 1986). From the work on alternate conceptions it has become obvious that students often construct alternate interpretations of the content from those that teachers expect. From the work on problem solving has come the realization that experts and novices have structured their knowledge very differently, and that the structuring of knowledge influences the manner in which each solves problems. Reif (1983), has drawn instructional implications from this research, especially in the area of physics, by developing models of expert and novice problem-solving performance, and teaching. Concurrently, our research group has studied related issues within transmission genetics with the goal being to improve genetics instruction at both the high school and college levels. The approach that we have taken has been to develop models of:

1. desired performance, by studying the problem-solving performance of geneticists and by analyzing the structure of transmission genetics (Collins, 1987; Collins & Stewart, 1987; Collins, Stewart & Slack, 1987);

2. novice performance, by studying college and high school students (Albright, 1987; Slack & Stewart, 1987; Stewart & Dale, 1981; Stewart, 1983); and
3. instruction in genetics (Streibel, Stewart, Koedinger, Collins & Jungck, 1987).

In this paper we report on a study of high school students' understanding of the physical relationship of chromosomes and genes, as expressed in their conceptual models, and in their ability to manipulate the models to explain solutions to dihybrid cross problems. We are particularly interested in how novices have structured their conceptual knowledge of genetics and meiosis and how it influences their problem-solving performance. First, the models of chromosome/gene organization and behavior that students used to explain their solutions to dihybrid cross problems will be discussed. This will be followed by a detailed analysis of explanations given by three students: one who used a nearly correct model to explain the results of a correct problem solution; one who used an incorrect model to explain the results of a correct problem solution; and one who used an incorrect model to explain the results of an incorrect solution. We shall argue that the manner in which these students manipulated their models does not differ in any significant way, and that both erroneous and correct models function to explain problem solutions. This leads to the conclusion that, in addition to the tendency to give too much credit to students who obtain the expected answers to problems (Stewart & Dale, 1981), there may be a tendency to underestimate the knowledge and abilities of students who do not obtain right answers.

The results of this research are currently being used to develop revised instruction on meiosis and genetics for high school students including an intelligent tutoring system for transmission genetics, MENDEL (Streibel et al., 1987).

Methods

Fifty high school students took part in this study. Half were from a city school with an enrollment of 1400 students in grades 9-12 and half from a rural/small city school with a 9-12 enrollment of 700. All were either freshmen or sophomores enrolled in an introductory biology course. There were 21 females and 29 males. The students from the smaller school made up one section of one teacher's biology course, while those from the larger school were a sample selected from the 100 students in four sections of a single teacher's biology course.

The students in each school received at least one month of instruction in meiosis, basic transmission genetics (simple dominance and codominance), advanced transmission genetics (multiple alleles and linkage), and molecular genetics. After completing the instruction on meiosis and basic transmission genetics, each of the 50 students took part in a 50-minute problem-solving/interview session. In it they each were asked to solve a monohybrid and a dihybrid problem (involving simple dominance) while thinking aloud. An example of a problem that was presented to students is:

> In beetles, wings with spots are dominant to wings without spots, and long antennae are dominant to short antennae. What are the possible offspring genotypes and phenotypes of a cross between one beetle that is heterozygous for wing spots and heterozygous for long antennae and a second beetle that has unspotted wings and is heterozygous for antennae length?

Following the solution to the dihybrid problem, students were asked a set of questions to elicit their models of chromosomes and genes and to describe their movement during meiosis. The interviewer asked an open-ended question such as "Do chromosomes have anything to do with your problem solution?" Each student who produced correct genotypes was also asked why he/she had not shown gametes around their Punnett square with two alleles of the same gene rather than one allele from each of the two genes. Although the subsequent question sequence was tailored to individual students, our purpose was to generate details about how students thought chromosomes and genes could be used to account for their solutions. Only technical language which a student first introduced was used in an interview. For example, the researcher did not call a structure a gene until after a student had first used the term. Once students acknowledged that chromosomes were involved in their solution to the dihybrid problem (and not all did) they were asked to indicate how many chromosomes were involved and to draw pictures of how the symbols that they had used in solving the problems could be shown in a diagram that also included chromosomes. They were then asked to explain how their models related to their solutions. This question usually led to their demonstrating how their models could account for the arrangement of symbols around their Punnett squares. In other words, they were asked to describe how they thought their models were related to meiosis, although the term meiosis was not used if the student didn't introduce it. In the course of this questioning, some students constructed, or were asked to consider, other models. Each of the interviews was recorded and transcribed. Everything that students wrote or drew during the interviews, as well as the interview transcripts, constituted the data that were analyzed.

Student Models

Of the 50 students interviewed, 41 obtained correct answers to both monohybrid and dihybrid problems. Thirty-five of those 41 were able to construct and discuss chromosome/gene models. Six students were able to solve monohybrid, but not dihybrid, problems. This group was evenly split in terms of whether or not they were able to construct and discuss chromosome/gene models. Finally, three students were unable to obtain correct solutions to either the monohybrid or the dihybrid problem. Of these three, only one was able to construct and discuss a chromosome/gene model. Those students who were unable to construct and discuss a chromosome/gene model either thought chromosomes had nothing to do with the problem solution or they thought that chromosomes and genes were involved and were unable to elaborate on the involvement.

The models which students constructed were either one-, two-, and four-chromosome models, with identifiable variants within each category. The total number of models per student is slightly more than the number of students who had models. This was because some students produced two models and were unable to choose between them or because they abandoned one model and constructed another in the course of their discussions. Such models were included in both categories.

It was clear from the students' transcripts that they used bits and pieces of genetics knowledge (not necessarily correct or well organized) to construct their models. The most important point was that most students knew that any model would have to account for the gametes around their Punnett squares. They also realized that genes and chromosomes are related, and most recognized that genes are on chromosomes (although some had chromosomes on genes or even traits on chromosomes). In addition, many students had incidental knowledge about chromosome doubling, splitting and crossing over that was used to get from the initial chromosome state to the chromosome state in the gametes. Students would often tinker with their models in order to bring them into agreement with the initial and gamete states.

One-Chromosome Models

Of the 39 students who developed models, nine produced and used a one-chromosome model. The three types of one-chromosome models are shown in Figure 1. The interesting thing about these models is that they work, at least in terms of producing gamete types that match the ones around the students' Punnett squares, even though not one of the three types shows any chromosomes doubling. Types #1 and #2 each involve the dividing of chromosomes, and the students who produced these models talked about the division (or splitting) of chromosomes. Students who produced model #2 even had a double division. Notice that there are two different approaches for insuring that, independent of the number of divisions, the gametes are of the correct type. Students who used models #1 and #2 followed their chromosome division(s) with an "all possible combinations" of the chromosome fragments approach. Students using model #3 began with the alleles arranged on the chromosomes in such a way that all possible combinations of gametes was assured from the start. Model #3 is the most unusual in that the gametes are already formed on the chromosomes and simply split off to go around the Punnett square.

Figure 1 Goes Here

Even though the students who used these one-chromosome models were able to account for their arrangement of gametes around a Punnett square, it is obvious that these models are more mathematical (e.g., involve an "all possible combinations" approach) than they are genetic. No student who used a one-chromosome model demonstrated much knowledge of the mechanism of meiosis.

Two-Chromosome Models

Of the 39 students who produced models, 19 produced six variations of a two-chromosome model. Four of those models are shown in Figures 2 and 3. As was the case with the one-chromosome models, each of these explains the students' gamete arrangements around their Punnett squares. Again, the students who produced these models had bits and pieces of knowledge about chromosomes, genes, gametes, and meiosis, and they also began with an end state, the organization of the gametes around their Punnett squares, and worked with their initial arrangement of chromosomes and genes so that they could account for the gamete types. While these two-chromosome models tend to have more indications of mechanism (e.g., chromosome doubling, division and crossing over) than do the one-chromosome models, most of them are still driven by an "all possible combinations" approach (see Figure 3 #3 and #4) rather than one fully derived from the concepts of genetics. Model #2, although not an "all possible combinations" model, is like an approach taken in the one-chromosome models -- all possible combinations of gamete types are ensured from the beginning, since one allele for each trait is located on the same chromosome and because there are different possible initial arrangements of those alleles. It also can be seen from examples #2, and #3 that there is a tendency to confuse chromatid and chromosome. Even though the students did not mention chromatids, the presence in their diagrams of what appear to be centromeres points to this confusion.

Figures 2 & 3 Here

One common feature of all two-chromosome models, which matches an event of meiosis, is that at some point the chromosomes split. The students seemed to be sure of this and often indicated that they could remember seeing such pictures in their texts or could remember their teachers making such drawings. Interestingly, some weren't thinking about splitting a doubled chromosome (or joined chromatids) into single (or daughter) chromosomes. For example, in model #2 there is a longitudinal split within a single chromosome that is more like the "unzipping" of DNA during replication than it is of chromosomal division (in fact one student who produced this model referred to the DNA splitting). Possibly these students were trying to recall particular things that they had read or heard. This was especially clear when they placed two allele symbols (either from the same or opposite gene pairs) on the same chromosome -- they could remember seeing pictures of single chromosomes containing more than one gene each. It was also logical to them that this should be the case since they knew that an organism has more traits than it does chromosomes. This searching for a "picture" was also true of one of the two students who invoked crossing over in his models -- he remembered seeing a picture of crossing over. (It is interesting that the other student apparently invented crossing over on his own, as it had not been studied in class before the interview.) Those models in which crossing over was used (#1 for example) involve a stronger use of mechanism than the other models, although even those do not deal with the importance of the arrangement of the chromosomes as they move from a doubled state through the two divisions of meiosis.

Four-Chromosome Models

Of the 39 students who produced models, 14 produced four-chromosome models. It should be pointed out that some of these students struggled with one- or two-chromosome models first. Of the 14 students who chose four-chromosome models, three obtained correct gamete types and three had essentially correct models. The major types of four-chromosome models are shown in Figures 4 & 5.

Figures 4 & 5 Here

A common feature of many of the students' four-chromosome models is that, like the one- and two-chromosome models, there is a tendency toward being mathematical rather than genetic. The students knew how to determine gamete types to use around their Punnett squares which they then used as a check on their chromosome/gene models. Thus they worked backwards from a known (gamete types) to a model. This led many of them to produce models that demonstrated little or no recognition of homologous chromosomes and little recognition that there is a mechanism that is responsible for the aligment of homologs in such a way that the correct gamete types are produced. For example, model #1 in Figure 4 is an example of the "all possible combination" approach. Although very confused about the spatial relationship of chromosomes, genes, and traits, this student nonetheless had two divisions that led to four traits (really alleles) that are separated. The student then took all possible combinations to produce the gametes. There is little sense of mechanism, only that there is a need to obtain all possible gamete combinations.

Although most of the 39 students who produced chromosome/gene models to account for their problem solutions were able to obtain correct answers to dihybrid problems, few were subsequently able to justify these correct answers by drawing upon a correct model of chromosome/gene behavior during meiosis.

Detailed Analysis of Three Students

In this section, the chromosome models used by three students will be examined in detail, with particular attention paid to the manner in which each student manipulated a model to explain or justify the procedural solution to a dihybrid problem. The students chosen are representative of the fifty students who took part in this study:

STUDENT A:	Incorrectly solved a simple dihybrid problem, and constructed an incorrect model to explain the incorrect procedural steps
STUDENT B:	Correctly solved a simple dihybrid problem and constructed an incorrect model to explain the correct procedural steps
STUDENT C:	Correctly solved a simple dihybrid problem and constructed a nearly correct chromosome model to explain the correct solution.

Each of these students' chromosome models and manipulations will be discussed separately, but for comparison purposes each student's solution steps and chromosome models are shown in Figure 6). To simplify the diagrams, each model shows only one of the parental genotypes.

Figure 6 Here

Student A

This student failed to solve the dihybrid problem correctly. As shown in Figure 6, he constructed a two-chromosome model (for a single parental genotype) with the alleles for a single trait on the same chromosome. Before questioning this student about his model, the interviewer (denoted by I in the following transcript excerpts) asked him (denoted by SA for Student A) if his Punnett square could be set up with gametes having two alleles of the same gene rather than one allele from each of two genes.

SA: Then it would've messed this up, cause I wouldn't get both of the traits in there...

I: Is there anything that goes on that ensures that you're always going to get the kind of combinations that you show [his Punnett square, see Figure 6]...instead of the combinations that I showed [the two alleles for the same gene]...

SA: When they breed...like the capital D...from the mother or say the father and then the small d from the other one...

A cursory look at student A's Punnett square might lead one to think that he simply took two alleles for a single trait from one parent and placed them around the square as shown in Figure 7A. However, the above discussion shows that student A knew that each parent contributed a single allele for a particular trait. Through manipulations of particular features of his chromosome model, he was able to describe a "meiotic mechanism" that results in the allelic combinations shown around his Punnett square.

I: You seem to be saying that you got the capital D from one parent and the little d from the other parent...

SA: That's through meiosis...breaks them all apart...going back to the genes, you get one allele from each parent, that makes up the gene...

I: ...show me how this process of meiosis works... [see Figure 7B]

SA: ...way it starts out. And then it's gonna divide and then replicate, so you end up...like that. And then if it was like eggs, any three of these would have died, along the way somewhere and you'd only end up with one.

I: So these [pointing to final 4 products] represent either the eggs or the sperm?

SA: Yeah.

I: Now is the same thing going to happen for the capital B and little b too?

SA: Yeah.

Figure 7 Here

Further questioning revealed that the student's allelic combinations were the result of this mechanism occurring in both parents, and that combinations such as Dd or Bb involved each parent contributing a single allele. Close examination of Student A's discussion and drawings revealed many gaps in his explanation of his procedural solution, and errors in his manipulation of a constructed chromosome model to fit the solutions steps he took. For instance, he constructed a chromosome model in which a pair of alleles for the same trait are linked. Also, he diagramed a chromosome "breaking apart" (the pair of alleles are separated), then doubling and finally "breaking apart" again to produce four gametes. Not surprisingly, given his chromosome model, he did not show any understanding of homologous chromosomes. Toward the end of the interview, Student A also considered a one-chromosome model and indicated that, with respect to his meiotic mechanism and this problem, it did not make much difference which model is correct. In spite of his errors and omissions, however, it would be a mistake to assume from his incorrect solution to a simple dihybrid problem that he had no understanding of genetics. He knew that each parent contributed a single allele for each trait and that offspring receive a full complement of genes. Furthermore, he under-stood that his procedural solution to this dihybrid problem is grounded in real world objects (chromosomes, eggs and sperm) and events (meiotic division).

Student B

This student had no difficulty producing a correct solution to a simple dihybrid problem, but developed an incorrect two-chromosome model to explain his solution steps. He was asked to circle anything that represented a gamete in the problem solution, and he correctly identified several. Then,

I: How do you know that's a gamete?

SB: When it goes through meiosis, these are the possible.-..gametes that can be formed...

I: When what goes through meiosis?

SB: When um, a chromosome. Well that's what the traits are on. Traits are on chromosomes.

[Student B then produced the diagrams shown in Figure 8A.]

I: How many chromosomes have you shown me here?

SB: Four. But really two but, um, they replicate...they replicate and then they form with their homologous, homologous pairs. And they go through division.

The student then produced the diagram in Figure 8B, and explained it by saying:

> ...you have the chromosomes......then what they do is replicate, and so you have a, oh, another D over here and another big B...little b and little d and then...when they go through division, then you got, um...four gametes with different

> combinations...And they could be different, just depending on crossovers and um, it, depending on crossovers through the chromosomes.

Figure 8 Here

It is clear that his chromosomal manipulation diagrammed thus far will result in only two of the four gamete types shown in the problem's solution. The possible "solution" to this difficulty is hinted at in the last part of the excerpt, and when the interviewer explicitly asked how this model could account for the other two gamete types, the student produced the diagram shown in Figure 9 and explained:

> Um, well during meiosis, um, think it's during...either metaphase or anaphase, one of those two. Um, the chromosomes will, sort of break off. And will, let's see. It'll end up like that. Let me see. And that part goes over to this part, and this goes...and so what you have is um, like you have, one...one, of the regular chromosomes and one part, of the broken off chromosome...then you have, that part and then you have this part of the broken off chromosome, so they could, mix. So they, that's how you get, you know, different combinations of different, um, gametes...

Thus each parental genotype represents two chromosomes with an allele for each trait on each chromosome. All possible gamete types are explained through a crossing over mechanism that takes place during meiosis. It is interesting to note that when student B was asked what, if anything, prevented the formation of gametes such as Aa, and Bb, his initial response was to state that it would "...end up with a lot of problems if you have this kinda combo...you can't really...it'd be a mutation." When questioned further on this issue, however, he was able to talk about "homologous chromosomes lining up". Within the context of his chromosome model the homologs are shown in Figure 8D. When presented with an alternative one-chromosome model, Student B simply stated that the chromosome "...couldn't divide if they [the traits] were on the same chromosome."

Figure 9 Here

Student C

This student correctly solved the dihybrid problem and constructed a chromosome model that demonstrated a reasonable understanding of the genetics underlying his correct procedural steps. This student is also distinguished by an increased ability to consider alternative chromosome models and reject incorrect ones. He does lack an important concept -- the pairing of homologous chromosomes during meiosis-- the absence of which makes his explanations incomplete and his acceptance of a four-chromosome model somewhat tentative.

Student C correctly identified the pairs of symbols on the outside of the Punnett square as gametes. When asked why he had not constructed pairs such as Bb and Dd, he replied,

> ...they are supposed to be gametes. And they've only got one of a kind. Gametes should only have, either one part of the trait or the other, the dominant or recessive...No, it can't have two of these. This couldn't be a gamete.

When asked if anything insures that such pairings will not occur in organisms, he identified meiosis. More importantly, he was able to diagram a meiotic division, initially utilizing just the symbols representing the genotype of a single parent (chromosomes were not discussed at this point). He elaborated on this representation in the following exchange.

> I: Would it be possible for you to take, this makeup of the parent, (BbDd) through meiosis?
>
> SC: OK...Put 'em all together first...The cell. And they replicate...I'm also gonna put another one in here, so I don't hafta draw another little one.
>
> I: OK.
>
> SC: This is where they get together in little, groups of four. [pause] I'll just draw two cells then. These become, two cells when it splits down like this, which means...like that, and like that. It could be, the other way around. You don't hafta have little b's on one side.
>
> I: I see.
>
> SC: And the, doesn't, and then it divides right away again, so we'll divide down like that, and divide it, like that. To get...four little ones with, d, b and D, B. [pause]

I: And then these other...I mean you had two different, types there,

SC: Yeah.

I: and then this, you had four, and that would come about- ...and these other two you could account for how then? Like the little d and large B and large D and little b?...

SC: It's probably just the way they're lined up because I, they don't hafta be in...any real particular order, uh, these could be over here.

The manipulation of allelic symbols through a meiotic division shows a reasonable understanding of the genetic mechanisms that explain his procedural solution steps. When asked if (and then how many) chromosomes are involved in this problem solution, student C's initial response was that there were two chromosomes for each parental genotype. His model is shown in Figure 10A, and a meiotic division of it is shown in Figure 10B.

Figure 10 Here

After drawing his model and meiotic division, he stated that the addition of chromosomes has complicated the matter:

> "I forgot about replication...I shoulda drawn four in the first place. There aren't two chromosomes, there are four!"

He then constructed a four-chromosome model as shown in Figure 10C.

He identified what he considered to be a major problem with this model, however:

> they'd run into problems, depending on how they lined up. If you kept 'em lined up exactly like this they'd still end up with two of this trait here and none of this other one.

What was absent from his otherwise reasonable model and explanation is the idea that homologous chromosomes pair during meiosis, ensuring that the "wrong gametes" are not produced. The absence of this concept not only made his explanation incomplete, it led him to consider a two-chromosome model (Figure 10D) as plausible. His reasoning was that such a model solved the problem he had with his four-chromosome model. In the end he returned to a four-chromosome model because he recognized a different problem with his two-chromosome model: how would such a model produce all four gamete types shown around the Punnett square?

Discussion of the Three Students

Obtaining right answers to dihybrid genetics problems is often simply a function of having the appropriate algorithms(s) at one's disposal. If arriving at correct answers were the sole indicator for judging success and learning following genetics instruction, then we would have no difficulty discriminating Student A from Students B and C. However, if we are also interested in ascertaining whether or not students are "thinking genetically"--where such thinking is evidenced at least in part by the ability to explain (in terms of the underlying chromosomal and meiotic mechanisms) the algorithmic solution steps taken--then Students A, B, and C cannot be so simply characterized and categorized.

Each of these students used concepts of genetics in solving dihybrid problems. This was evidenced by their ability to: (1) construct a chromosome model, and (2) manipulate this model with an explicit reference to a process each identified as meiosis, in order to explain the prior procedural solution. This is not to say that no differences existed among the three students. Student C's model and explanation were more nearly correct than either of the others, and Student C was the only one to consider and reasonably rule out alternative incorrect models (although it should be remembered that because Student C did not seem to have a working knowledge of homologous chromosomes, a two-chromosome model represented a tempting alternative).

Students A and B make what could be described as interesting errors. Despite the gross procedural errors characterizing Student A's solution to the dihybrid problem, his explanation of that solution clearly indicates some knowledge and understanding of the underlying genetics and a demonstrated ability to apply that knowledge in a reasoned manner. The same can be said about Student B. Their mistakes, as well as those of Student C, are interesting and useful when viewed within a pedagogical context. If they are not completely idiosyncratic (and the three types of chromosome models described in this paper would indicate that they are not), they can point out areas to which genetics instruction must be sensitive, such as homologous chromosomes, allelic linkage, and crossing-over.

Discussion and Implications

There are a number of implications that can be drawn from our research on high school students' understanding of chromosome/gene behavior. We take it as a given that a central focus of genetics instruction is that students will develop understandings of causal mechanisms such as meiosis which underlie and give meaning to problem solving. Without an understanding of underlying mechanisms it is much more likely that students' solutions to genetics problems will be algorithmic (Stewart & Dale, 1981; Stewart, 1983). It is also the case that without understanding at the level of mechanism it is difficult, if not impossible, to solve more realistic genetics problems (Stewart, 1987). Elsewhere we (Collins, Stewart, and Slack, 1987) have reported that PhD geneticists, when facing challenging realistic problems, use underlying mechanisms such as meiosis to construct hypotheses that allow them to work towards a solution.

Given the above, and the results of our research that suggest that high school students may have numerous alternative views of meiosis following instruction, we feel that our research has the following implications for those interested in ensuring that high school students develop an accurate understanding of meiosis.

1. Teachers need to be aware that alternate views, such as the ones described in this paper, may be common outcomes of instruction. Therefore, they need to take steps to reduce the likelihood that these alternative views will occur, and to identify them if they do occur.

For example, most of our students, no matter what model they used, recognized that chromosomes double and divide (or split). It seem obvious then, that students have attended to some details of the instruction but haven't always learned more important concepts. Teachers need to be careful when teaching about meiosis that students understand the most salient features of the process. This may mean leaving out much of the detail that is currently found in genetics instruction (see Thomson & Stewart, 1985).

2. When evaluating student problem solutions teachers should look at more than the answer. Students should be expected to justify answers at whatever level of mechanism they have been taught. Persuasion plays an important role in science. Scientists are constantly in the position of having to persuade their peers that the results of their research (problem solving) is logical and that there is a consistency between theory, data, and claims. Students should be expected to do the same when solving problems. Not only will teachers gain insight into student thinking, but students may develop more meaningful understanding of genetics and of what it means to "do" science.

3. In an earlier paper (Stewart and Dale, 1981) we argued that, since many students are able to obtain correct answers to genetics problems with minimal conceptual understanding of genetics, teachers must be careful that they do not give students too much credit for obtaining right answers.

In summarizing the results of this more detailed analysis of student understanding of meiosis, we are able to reiterate this claim. However, it is clear that such a claim does not tell the whole story. Many of the students who obtained correct answers, and even some who obtained wrong answers, did so by manipulating alternate chromosome/gene models in ways that accounted for their solution procedures. Therefore, we are concerned that current evaluation procedures only reward students for correct answers, and not the process of obtaining answers. This practice leads to rewarding some students who have little understanding and does not reward other students for imaginative model building because the outcome of the model building process is not a correct answer.

4. Since the students in our study were in either the ninth or tenth grade, it is reasonable to expect that many of them would probably have difficulty with combinatorial reasoning. Yet, very few students in this study had difficulty with the logic of combinatorial reasoning (in fact they were very inventive). Their difficulty was caused by a lack of conceptual knowledge necessary to manipulate the combinations in correct ways. Students whose models were "wrong" nonetheless manipulated them using sophisticated combinatorial reasoning.

References

Albright, W. (1987). A description of the performance of university students solving realistic genetics problems. Unpublished MS Thesis. University of Wisconsin-Madison.

Chi, M., Feltovich, P., and Glaser, R. (1981) Categorization and representation of physics problems by experts and novices. Cognitive Science, 5, 121-152.

Collins, A. (1987) A description of the problem-solving strategies of experts solving genetics problems. Paper presented at the second International Seminar on Misconceptions and Educational Strategies in Science and Mathematics. Cornell University, Ithaca, New York.

Collins, A., & Stewart, J. (1987). A description of the strategic knowledge of experts solving realistic genetics problems. MENDEL Research Report #1. Madison: The University of Wisconsin Department of Curriculum & Instruction.

Collins, A., Stewart, J. & Slack, S. (1987). Expert and novice problem solving in transmission genetics. Unpublished manuscript.

Driver, R. and Easley, J. (1978). Pupils and paradigms: A review of literature related to concept development in adolescent science students. Studies in Science Education, 5, 61-84.

Erickson, G. (1979). Children's conceptions of heat and temperature. Science Education, 53 (2), 221-230.

Hackling, M.W., & Treagust, D.F. (1982). What lower secondary students should understand about the mechanisms of inheritance and what they do understand following instruction. Research in Science Education, 12, 78-88.

Hackling, M.W., & Treagust, D. (1984). Research data necessary for meaningful review of grade ten high school genetics curricula. Journal of Research in Science Teaching, 21 (2), 197-209.

Kargbo, D.B., Hobbs, E.D., & Erickson, G.L. (1980). Children's beliefs about inherited characteristics. Journal of Biological Education, 14 (2), 137-146.

Larkin, J. (1983). The role of problem representation in physics. In D. Gentner & A.L. Stevens (Eds), Mental Models. Hillsdale, N.J.: Lawrence Erlbaum Associates.

Reif, F. (1983) Understanding and teaching problem solving in physics. Research on Physics education: Proceedings of the first international workshop. Paris: Editions du Centre National de la Reserche Scientificue.

Slack, S. & Stewart, J. (1987) Student Problem Solving in genetics. Paper presented at the second international seminar on misconceptions and educational strategies in science and mathematics. Cornell University, Ithaca, New York.

Smith, M.U., & Good, R. (1984) Problem solving and classical genetics: Successful vs. unsuccessful performance. Journal for Research in Science Teaching, 21 (9), 895-912.

Stewart, J. (1987). Potential Learning Outcomes from Solving Genetics Problems: A typology of problems. (In Press) Science Education.

Stewart, J. (1983). Student problem solving in high school genetics. Science Education, 67 (4), 523-540.

Stewart, J. and Dale, M. (1981) Solutions to genetics problems: Are they the same as correct answers? The Australian Science Teacher, 2713: 59-64.

Streibel, J., Stewart, J., Koedinger, K., Collins, A., and Jungck, J. (1987) MENDEL: An Intelligent Computer Tutoring System for Genetics Problem Solving, Conjecturing, and Understanding. Machine Mediated Learning 2 (1 & 2) 129-159.

Thomson, N. & Stewart, J. (1985). Secondary school genetics: Making problem-solving instruction explicit and meaningful. Journal of Biological Education, 1985, 19, 53-62.

Acknowledgements

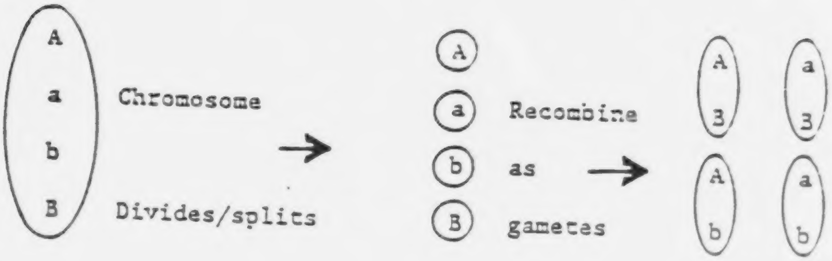
This article was prepared with support from the National Science Foundation under Grant No. SED 8022912. Any opinions, findings, conclusions, or recommendations expressed in this article are those of the authors and do not necessarily reflect the views of the National Science Foundation.

The authors also wish to acknowledge the helpful comments of Peter Hewson, Norman Thomson, and Judith Van Kirk on earlier drafts of this paper.

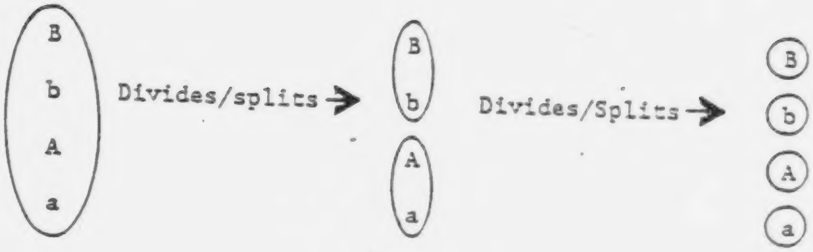
Figure 1

One-Chromosome Models Constructed by Students

1.


A
a
b
B
Chromosome
Divides/splits
A
a
b
B
Recombine
as
gametes
A
B
a
B
A
b
a
b

2.


B
b
A
a
Divides/splits
B
b
A
a
Divides/Splits
B
b
A
a

Then take all possible combinations of A's and B's.

3.

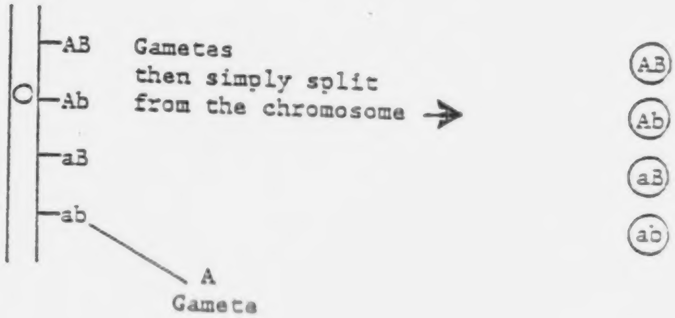
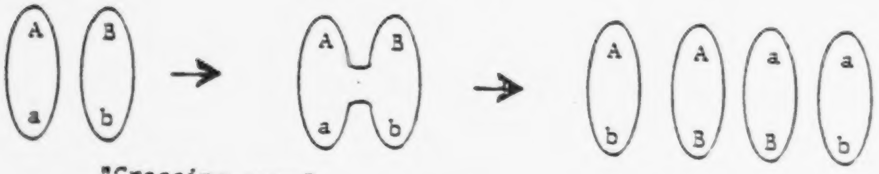

AB
Ab
aB
ab
C
Gametes then simply split from the chromosome
A Gamete
AB
Ab
aB
ab

Figure 2

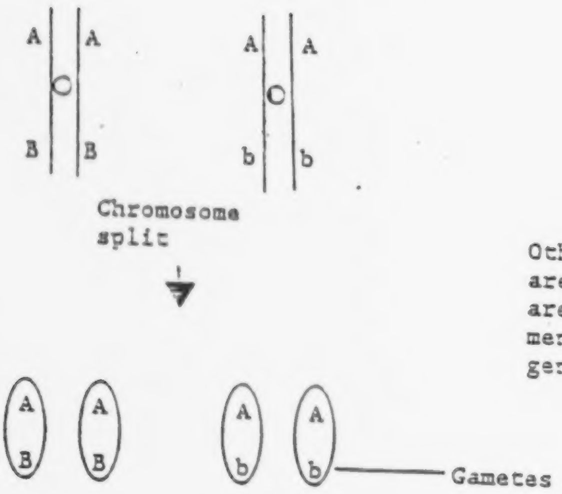
Two-Chromosome Models Constructed by Students

1.


A
a
B
b
A
B
a
b
A
b
A
B
a
B
a
b

"Crossing over"
1 Trait\gene on each chromosome

2.


A
B
A
B
A
b
A
b
A
B
A
B
A
b
A
b
Gametes

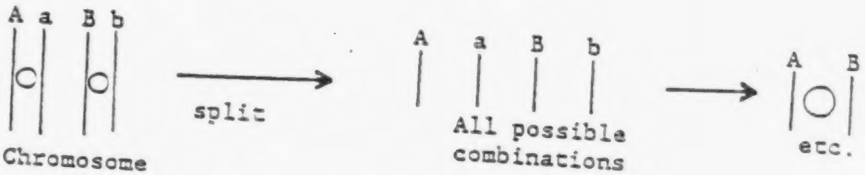
Chromosome split

Other gamete combinations are possible because there are other initial arrangements of chromosomes and genes.

Figure 3

"All Possible Combinations"
Two-Chromosome Models Constructed by Students

3.


A a B b
Chromosome
split
A a B b
All possible combinations
A B
etc.

4.

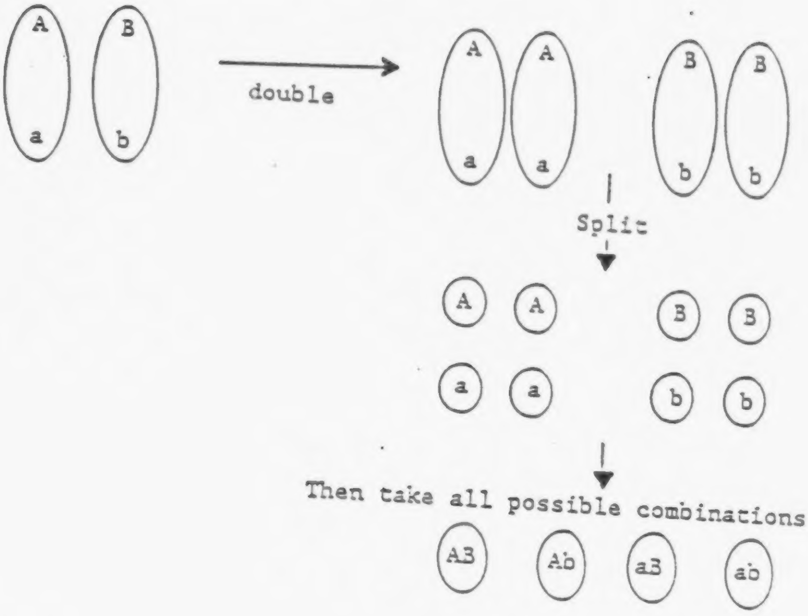
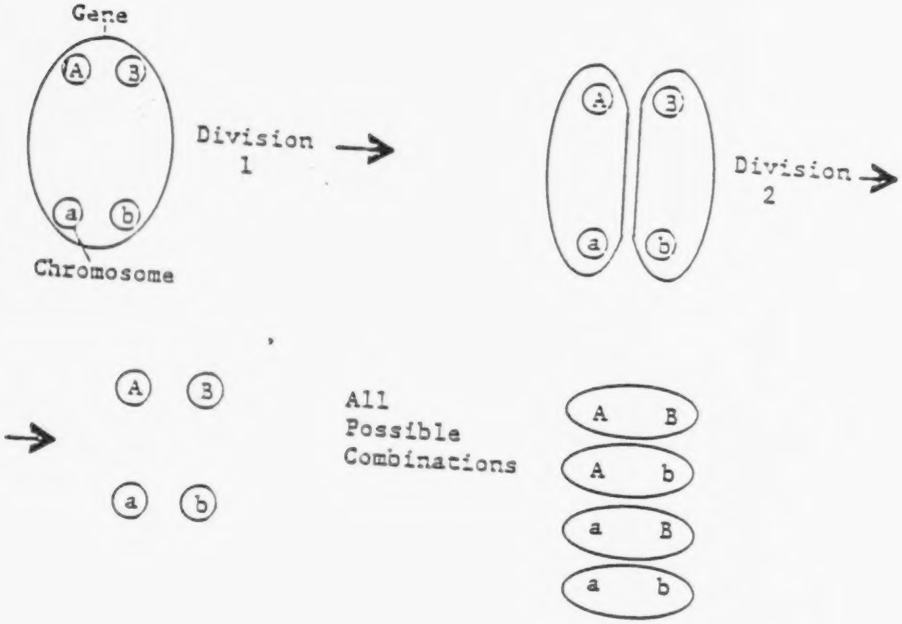

A
a
B
b
double
A
a
A
a
B
b
B
b
Split
A A B B
a a b b
Then take all possible combinations
AB Ab aB ab

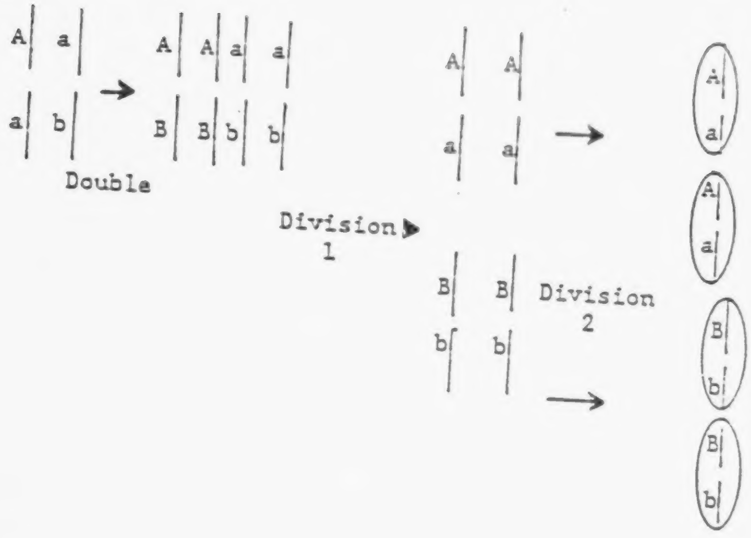
Figure 4

Four-Chromosome Models Constructed By Students

1.


Gene
A
B
a
b
Chromosome
Division 1
A
B
a
b
Division 2
A
B
a
b
All Possible Combinations
A B
A b
a B
a b

2.


A a
a b
Double
A A a a
B B b b
Division 1
A A
a a
B B
b b
Division 2
A a
A a
B b
B b

The student did not like these results so he created all possible correct combination

Figure 5

More Four Chromosome Models Constructed by Students

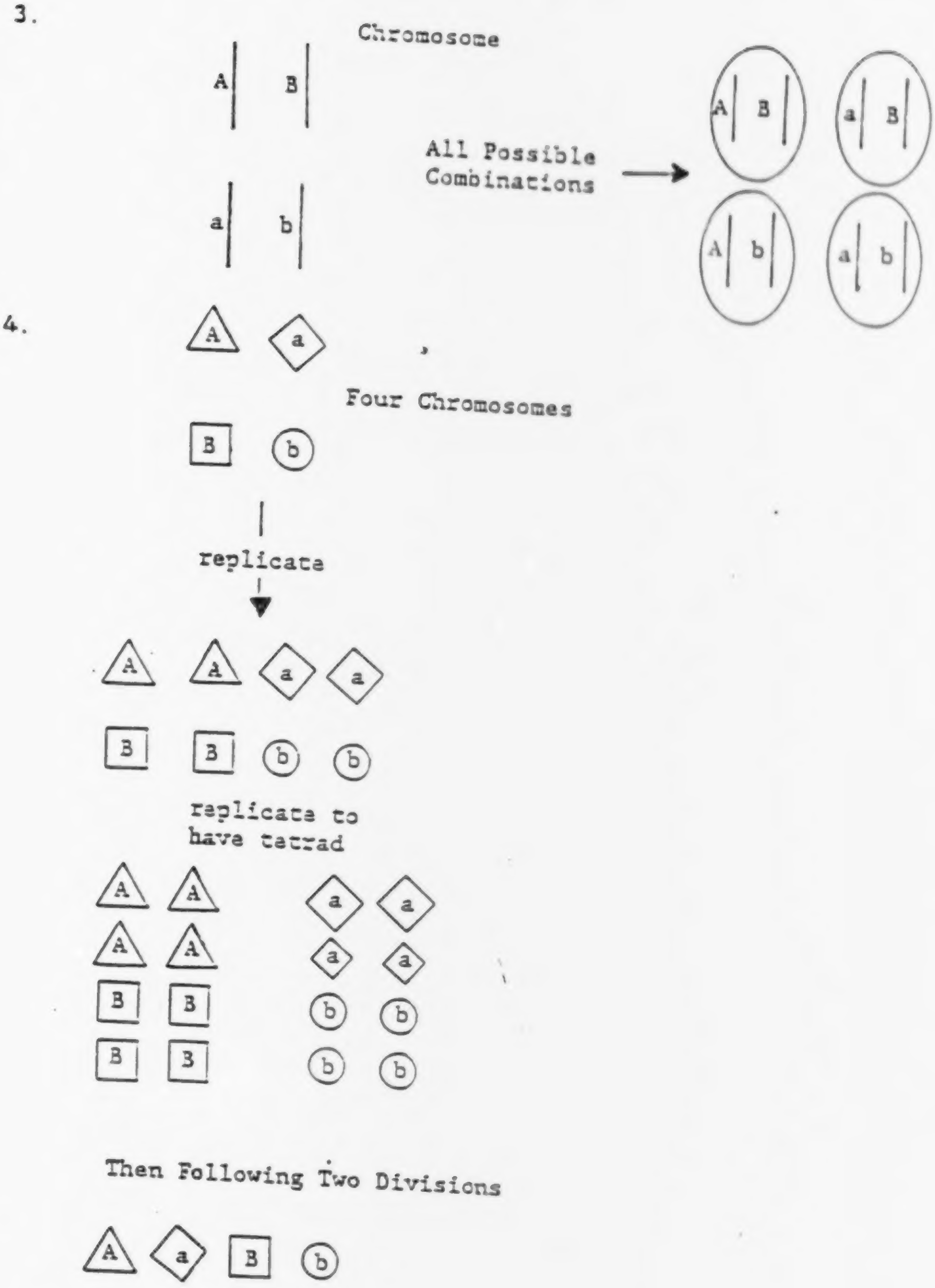

3.
Chromosome
A
B
a
b
All Possible Combinations
A B
a B
A b
a b
4.
A
a
Four Chromosomes
B
b
replicate
A A a a
B B b b
replicate to have tetrad
A A a a
A A a a
B B b b
B B b b
Then Following Two Divisions
A a B b

FIGURE 6

Data Relevant to the Discussion of Students A B and C

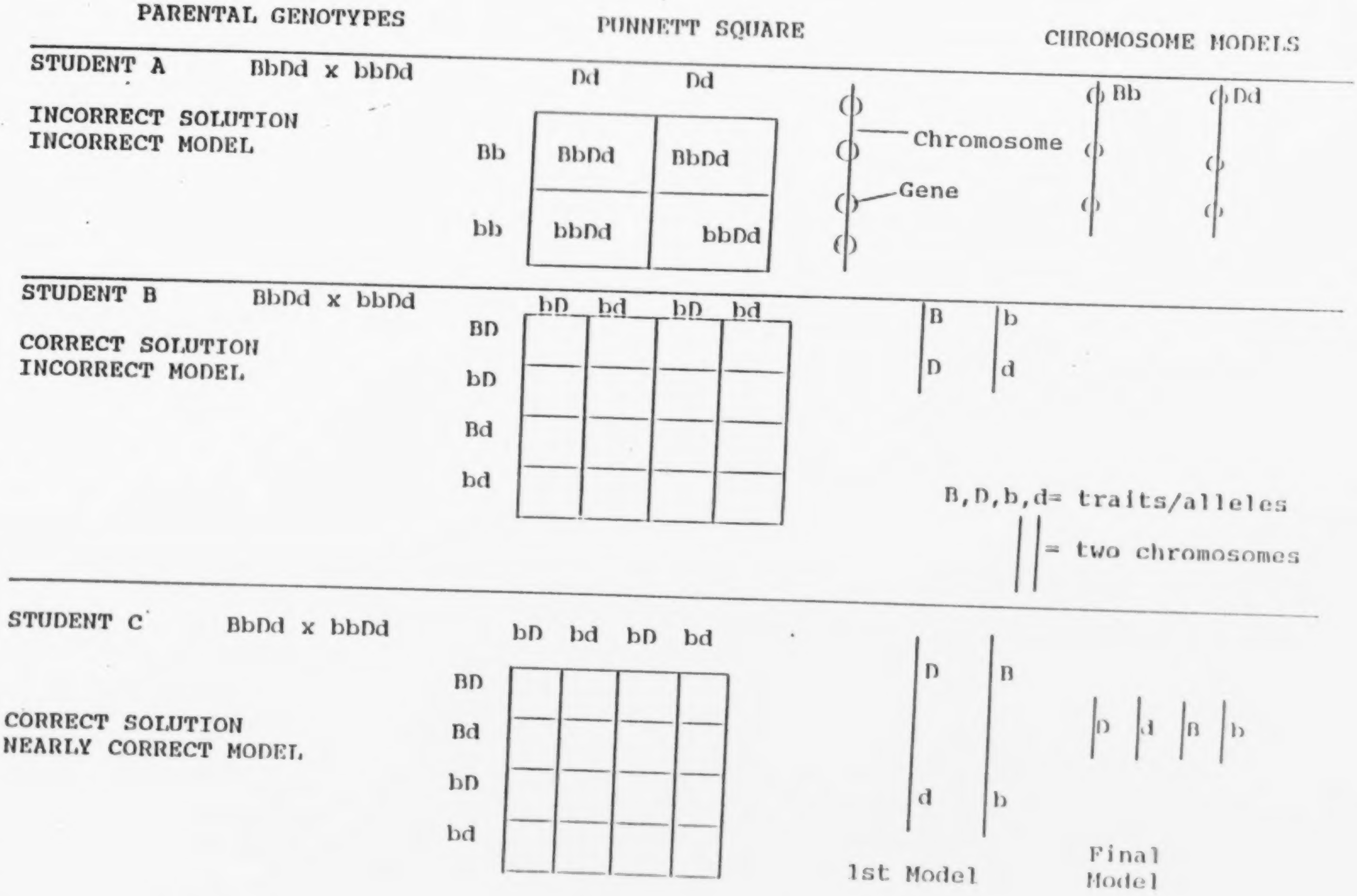

PARENTAL GENOTYPES
PUNNETT SQUARE
CHROMOSOME MODELS
STUDENT A
BbDd x bbDd
INCORRECT SOLUTION
INCORRECT MODEL
Dd
Dd
Bb
bb
BbDd
BbDd
bbDd
bbDd
Chromosome
Gene
Bb
Dd
STUDENT B
BbDd x bbDd
CORRECT SOLUTION
INCORRECT MODEL
bD
bd
bD
bd
BD
bD
Bd
bd
B
D
b
d
B,D,b,d= traits/alleles
= two chromosomes
STUDENT C
BbDd x bbDd
CORRECT SOLUTION
NEARLY CORRECT MODEL
bD
bd
bD
bd
BD
Bd
bD
bd
D
d
B
b
1st Model
D
d
B
b
Final Model

Figure 7

Student A's Punnett Square
and Chromosome/Gene Model

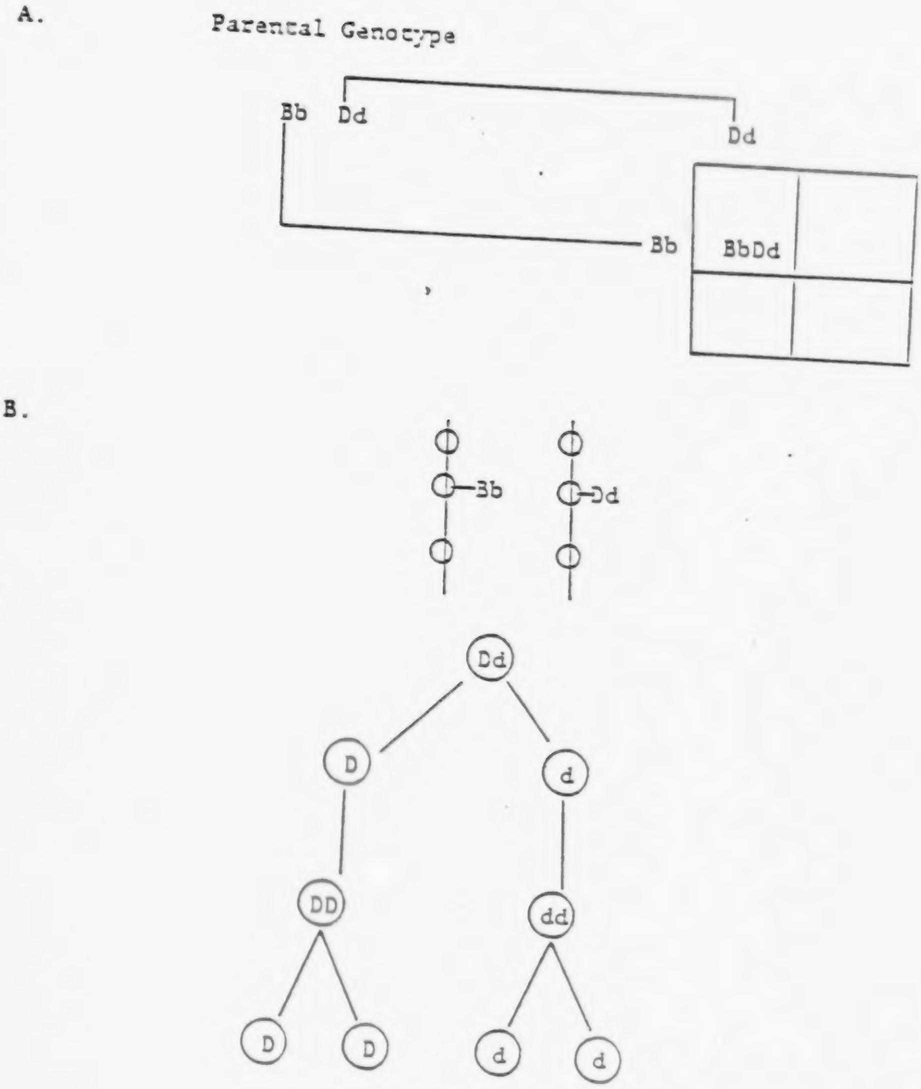
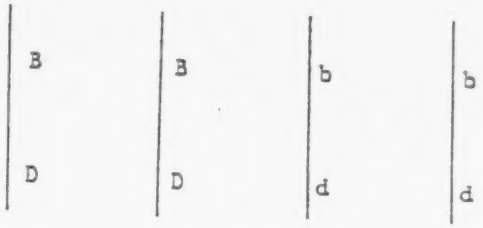

A.
Parental Genotype
Bb
Dd
Dd
Bb
BbDd
B.
Bb
Dd
Dd
D
d
DD
dd
D
D
d
d

Figure 8

Student B Chromosome/Trait Diagrams

A.



B.

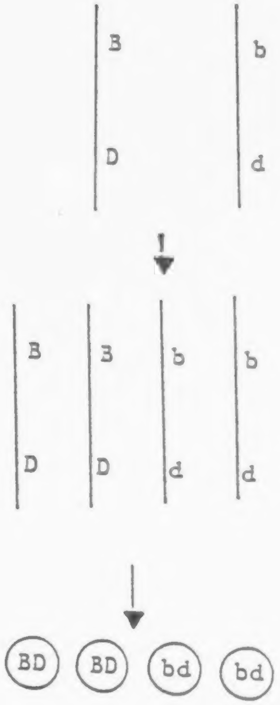
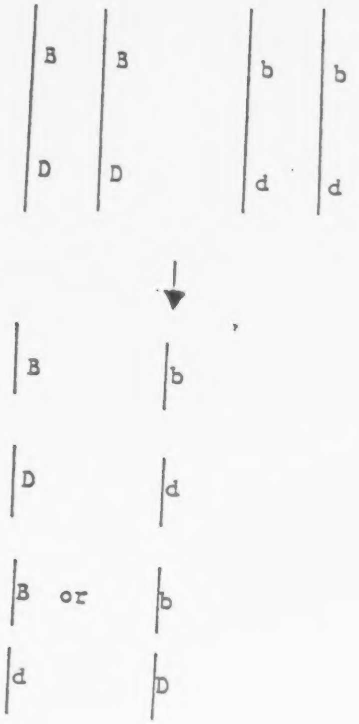

B
D
b
d
B
D
B
D
b
d
b
d
BD
BD
bd
bd

Figure 9

Student B's "Crossover" and "Homologs" Diagram

C.


B
B
b
b
D
D
d
d
B
b
D
d
B
or
b
d
D

D.

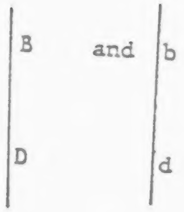
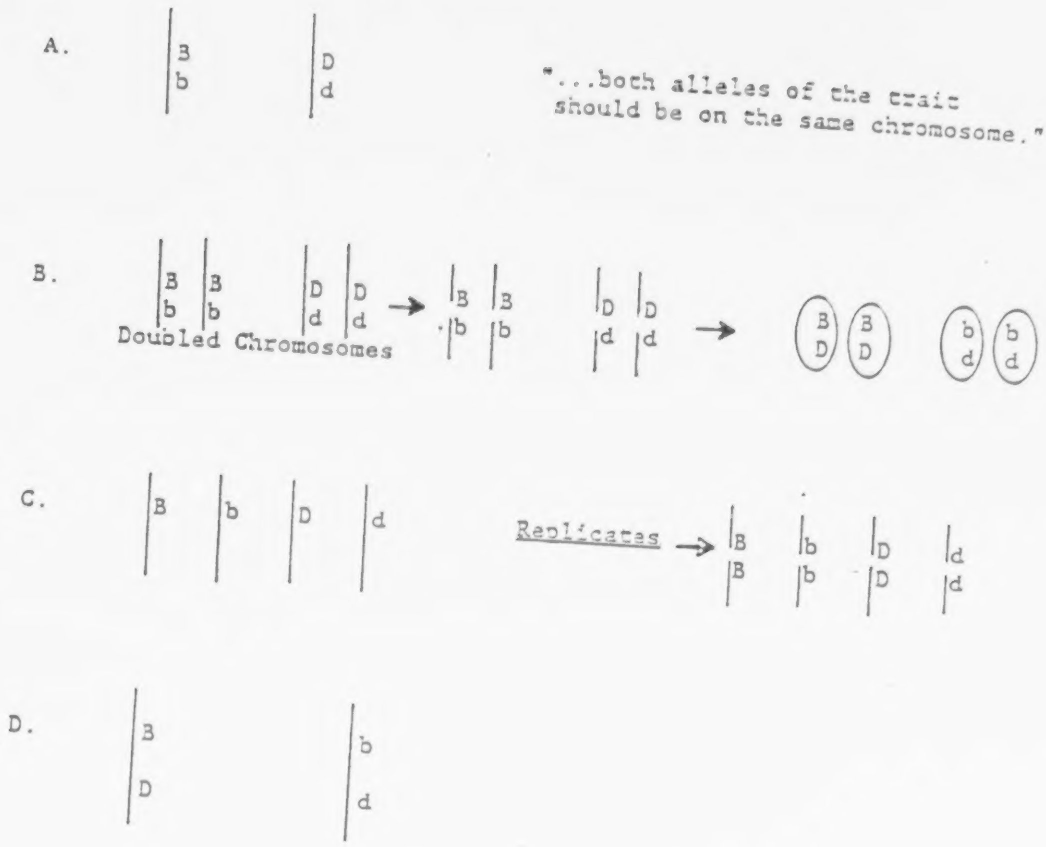


Figure 10

Student C Chromosome Gene Model


A.
B
b
D
d
"...both alleles of the trait should be on the same chromosome."
B.
B
b
B
b
D
d
D
d
Doubled Chromosomes
B
b
B
b
D
d
D
d
B
D
B
D
b
d
b
d
C.
B
b
D
d
Replicates
B
B
b
b
D
D
d
d
D.
B
D
b
d

LIST OF FIGURES

Figure 1: One-Chromosome Models Constructed by Students

Figure 2: Two-Chromosome Models Constructed by Students

Figure 3: "All Possible Combinations" Two-Chromosome Models Constructed by Students

Figure 4: Four-Chromosome Models Constructed By Students

Figure 5: More Four Chromosome Models Constructed by Students

Figure 6: Data Relevant to the Discussion of Students A B and C

Figure 7: Student A's Punnett Square and Chromosome/Gene Model

Figure 8: Student B Chromosome/Trait Diagrams

Figure 9: Student B's "Crossover" and "Homologs" Diagram

Figure 10: Student C Chromosome Gene Model

APPENDIX E

Definition of Terms Used in MENDEL

1. **Chromosomes**: A chromosome (more properly a chromosome pair) is seen as containing two important pieces of information:
    a. **Sex Linked**: if a chromosome is sex linked (noted by a Boolean true) then in the individuals that have the sex linked chromosome (for program purposes the male is always the bearer of sex linked chromosomes) the Y or sex linked half of the chromosome contributes no information (symbolically represented as NIL).
    b. **A list of loci**: a chromosome may be viewed as a linked list of loci, and our interest lies in a subset of these loci. The implied relation of the loci by their relation on a list will be discussed in the linkage topic.

2. **Loci**:
    a. **A locus**: a location on a chromosome (and inherits the sex linkage value).
    b. **A list of alleles** which may occur at that locus.

3. **Allele**:
    a. An **allele** is a name or tag given to a set of values that may occur at a locus.
    b. It is important to note therefore that we are working with discrete values as alleles. Discrete values allow us to completely define the problem-solving space.
    c. Since chromosomes are paired, a locus exhibits a pair of allele values with a sex-linked chromosome contributing only one value plus a NIL.

4. **Traits**:
    a. A **trait** is a physical manifestation of a locus or loci.
    b. A trait has a number of variations which depend on the genotypes of an individual at that particular locus or loci.
    c. The list of all possible genotypes for a trait and the variation that corresponds to that genotype is the Expression Chart for that particular trait.

5. **Genotypes**:
    a. A **genotype** is a list of sublists, each sublist representing a pair of allele values contributed by a locus.
    b. Example 1: If a trait is dependent on one locus whose possible allele values are A and B then the possible genotypes for this trait are ((AA)), ((AB)), and ((BB)).
    c. Example 2: Another trait dependent on two loci, whose possible allele values are A and B for one locus and C and D for the other has possible genotypes ((AA) (CC)),((AA) (CD)),((AA) (DD)),((AB) (CC)),((AB) (CD)),((AB) (DD)),((BB) (CC)),((BB) (CD)) and ((BB)

(DD)).

d. Generally we will be dealing with traits dependent on only one locus. In these cases the genotypes are referred to as being either homozygous (the genotype consists of a pair of the alleles with the same name e.g.((AA))), or heterozygous (the genotype consists of a pair of alleles with different names e.g.((AB))).

6. **Expression Charts**:
   a. A trait's **Expression Chart** consists of a list of sublists of the form: (Genotype Variation-Name).
   b. Example 1: using the possible genotypes from example 5.b.1 a possible expression chart might look like:
      ((((AA)) Large) (((AB)) Medium)(((BB)) Small))
   c. **Assumption 1**: (one locus only) genotypes can only represent the same variation when they share at least one allele (that is ((AA)) and ((AB)) could be the same variation (because they share the allele A), but ((AA)) and ((BB)) could not be the same variation).
      Corollary 1a: no two homozygote variations may look alike.
   d. **Assumption 2**: If two heterozygote genotypes represent the same variation in the expression chart, then the homozygote of the shared allele also looks like the two heterozygotes. (Example if ((AB)) and ((BC)) look alike then ((BB)) also looks the same.)

7. **Inheritance Pattern (IP)**: a general way of describing the Expression Chart of a trait.
   a. One locus IP's: we are generally interested in this small subset of IP's.
      i. two allele IP's: two allele IP's consist of two homozygote plus one heterozygote genotype.
         - Simple Dominance (SD): an inheritance pattern that has two variations where one of the homozygotes plus the heterozygote represent one of the variations and the other homozygote represents the other variation.
         - Codominance (CD): an inheritance pattern that has three variations where each variation is represented by one of the genotypes.
      ii. Multiple (more than 2, generally 3) alleles (MA) IP's: more general than either SD or CD.
         - In MA with 3 alleles there are 3 homozygotes and 3 heterozygotes.
   b. **Interaction** (more than 1 locus, generally 2 loci) IP's:
      i. These inheritance patterns represent a large percentage of the problem space, but because these patterns are complex even in the simplest of cases we are not as interested in them for tutoring purposes.

8. **Phenotype**:
   a. A property of an individual, a phenotype is a list of what variations a particular individual exhibits, one for each trait.
   b. The phenotype plus the sex of the individual is the information the student or the solver should be able to access. Information such as particular genotype is restricted to the GENERATOR.

9. **Modifiers**:

- a. **Pleiotropy**:
    - i. Two traits are pleiotropic if they are both dependent on at least one shared locus. If both traits have only one locus then both traits are dependent on that one locus.
- b. **Sex Linkage**:
    - i. Sex Linkage as a factor causes the genotypes of traits dependent on loci that are sex linked to be different for male and female individuals.
    - ii. Example 1: If in example 5.b.1 the traits' locus were on the sex linked chromosome the possible female genotypes would be unchanged but the male possible genotypes would be ((A NIL)) and ((B NIL)).
    - iii. Assumption 3: the male variation always looks like the female homozygote made with the same allele. (ie ((A NIL)) will always look like ((AA))).
- c. **Autosomal Linkage**:
    - i. If two traits have loci that are on the same chromosome the traits may be linked (if the loci are close enough together).
    - ii. Generally linkage means that the alleles from the two loci on one member of the chromosome pair are more likely to be passed together.
    - iii. The likelihood that both members are passed together is defined in terms of distance between the loci.
    - iv. Linkage may also occur between more than two loci.
- d. **Interference**:
    - i. Interference occurs in concert with Autosomal Linkage. There must be at least three loci on the same chromosome. Interference occurs when the distance between the two furthest apart loci is not functionally equal to the sum of the distance from each of these loci to the center locus.
- e. **Lethality**:
    - i. A genotype that is a lethal causes any individual that has that genotype to be dead.

APPENDIX F

Extended Definitions of Terms Used in MENDEL

1. **Population**:
   a. Chromosome list of the individuals.
   b. Trait list of the traits in the population we are interested in examining.
   c. Linkage matrix of the probability that the allele values from two loci will occur together. 0 represents that the probability is determined randomly.
   d. Offspring list in the population.

2. **Chromosome**:
   a. Sex Linked boolean: whether or not the chromosome is sex linked.
   b. Loci list: the chromosome contains.

3. **Locus**:
   a. Allele list of the possible allele values for the locus.
   b. Traits Affected list: what traits the locus influences.

4. **Trait**:
   a. Loci list of which loci determine the trait.
   b. Variations: what variations are possible in the population for this trait.
   c. Expression Chart: a list of what genotypes will correspond to which variations in the population.

5. **Offspring**: (each one represents a separate vial in the population)
   a. Vial-descriptor: description the physical vial.
   b. Sex Class list: of the offspring

6. **Sex Class**: one for each combination of different phenotypes plus sex in a population.
   a. Phenotype for the sex class as defined above.
   b. Sex of the Sex Class.
   c. No-of-Individuals in the Sex Class.
   d. Individual-Matrix: a matrix of the allele values an individual has for each locus in the population.

7. The definition of the characteristics of the initial population is referred to as **CUSTOMIZATION**.
   a. CUSTOMIZATION is implemented using a menu which contains probabilistic and definitive information concerning the population. Information as to which IP's and modifiers occur in the population and the probabilities of occurrence are set by the user(teacher). The information as to how many traits plus the minimum and maximum progeny are also set by the user(teacher). The user(teacher) therefore defines a class of problems without defining the actual parameters of the population. The program then chooses a problem representation within this class of problems.

8. The creation of the initial population in the GENERATOR is called **MAKING INITIAL POPULATION**.
   a. The main interest in forming the initial population is to create a set of individuals from which all of the factors influencing the population may be discovered, but which does not indicate the factors through initial ratios.
   b. The method used in this initialization has to do with generating all the patterns the student might see and choosing a random set of these (though it is unlikely that these individuals will be representative of a random sampling of the underlying factors.)

9. The creation of vials is referred to as **CROSSing**.
   a. The choice of parents in a cross is left either to the user(student) or the solver, the job of the generator is to simulate the meiosis process with the individuals specified to produce an offspring which conforms to a normal curve random sampling of the possible offspring.
   b. The number of individuals and their sex is random though number of individuals is dependent on min and max progeny and sex on 50%.
   c. For each expected individual the meiosis process is performed and the individual (if not dead) is added to the resultant offspring.
   d. The basic meiosis process involves only a random number generator plus the linkage matrix to determine which parental genotypes will be chosen to form the offspring individual.

APPENDIX G

A Sample Problem and the Logic of the SOLVER

The **Initial Population** for this problem is shown in vial-A below:

```
---------------vial-A--------------
|Parents: Initial Population       |
|Traits:  ANTENNAE THORAX          |
| SEX   #   ANTENNAE   THORAX      |
------------------------------------
|  f    5   MISSING    TETRALTERA  |
|  m    1   MISSING    TETRALTERA  |
|  f    3   MISSING    DOTTED      |
|  m    5   MISSING    DOTTED      |
|  f    7   MISSING    CHECKED     |
|  m    6   MISSING    CHECKED     |
|  f    7   CRINKLED   TETRALTERA  |
|  m    5   CRINKLED   TETRALTERA  |
|  f    5   CRINKLED   DOTTED      |
|  m    8   CRINKLED   DOTTED      |
|  f    4   CRINKLED   CHECKED     |
|  m    5   CRINKLED   CHECKED     |
------------------------------------
```

From the initial population we observe the following:

1. There are two traits.
2. The 1st trait, ANTENNAE, has two variations: Missing and Crinkled.
3. The 2nd trait, THORAX, has three variations: Tetraltera, Dotted and Checked.

Since no crosses have been done yet, the only possible cross plans are to **cross unlikes**, so we will make plans to cross:

1. For trait ANTENNAE:
    a. Crinkled and Missing parents
2. For trait THORAX
    a. Dotted and Checked parents
    b. Tetraltera and Checked parents
    c. Tetraltera and Dotted parents

Proceeding into the strategy **Experiment**, we decide to generate a hypothesis for each trait:

1. For ANTENNAE, we know that the trait has two variations (from above), therefore we make the hypothesis that ANTENNAE is the result of Simple Dominance. Within this hypothesis, we are further able to determine that there are two possible expression charts:
    a. Missing is dominant:
        AA - Missing
        AB - Missing
        BB - Crinkled
    b. Crinkled is dominant:
        AA - Crinkled
        AB - Crinkled

BB - Missing

2. For THORAX, we know that the trait has three variations (from above), therefore we make the hypothesis that THORAX is the result of Simple Dominance. Within this hypothesis we are further able to determine that there are three possible expression charts:
   a. Tetraltera is the heterozygote:
      AA - Dotted
      AB - Tetraltera
      BB - Checked
   b. Dotted is the heterozygote:
      AA - Tetraltera
      AB - Dotted
      BB - Checked
   c. Checked is the heterozygote:
      AA - Tetraltera
      AB - Checked
      BB - Dotted

Proceeding into the **Test and Refine Hypothesis** strategy, we **Pick a Cross**, and decide to cross a Crinkled Dotted individual with a Missing Checked individual.

Results of Cross 1:

```
 ---------------vial-B--------------
|Parents: m vA CD i1  f vA MC i2    | Male parent vial A:
|Traits: ANTENNAE THORAX            |  - phenotype:
| SEX   #   ANTENNAE   THORAX       |     Crinkled Dotted
 -----------------------------------   - individual number 1
|  f    9   MISSING    DOTTED       | Female parent vial A:
|  m    6   MISSING    DOTTED       |  - phenotype:
|  f    7   MISSING    CHECKED      |     Missing Checked
|  m    6   MISSING    CHECKED      |  - individual number 2
|  f    5   CRINKLED   DOTTED       |
|  m    7   CRINKLED   DOTTED       |
|  f    4   CRINKLED   CHECKED      |
|  m    8   CRINKLED   CHECKED      |
 -----------------------------------
```

Examining the cross data from the first cross, we note that:

1. For trait ANTENNAE: Crossing a Crinkled male from vial A with a Missing female from vial A resulted in 15 Crinkled males, 9 Crinkled females, 12 Missing males and 16 Missing females.
2. For trait THORAX: Crossing a Dotted male from vial A with a Checked female from vial A resulted in 13 Dotted males, 14 Dotted females, 14 Checked males and 11 Checked females.

From this **Redescription**, we make cross plans:

1. For trait ANTENNAE:
   a. Cross a Missing male from vial B with a Missing female from vial B.
   b. Cross a Crinkled male from vial B with a Crinkled female from vial B.

   We also remove the plan to cross a Missing parent with a Crinkled parent since this cross was just done.

2. For trait THORAX:
    a. Cross a Checked male from vial B with a Checked female from vial B
    b. Cross a Checked male from vial B with a Checked female from vial B

   We also remove the plan to cross a Dotted parent with a Checked parent since this cross was just done.

Examining our hypotheses we determine that:

1. For trait ANTENNAE: Both hypothesized expression charts could explain the data. For example, in the expression chart where Missing is dominant the father could have been genotype BB, the mother genotype AB giving Missing and Crinkled offspring of both sexes with a 1:1 ratio.
2. For trait THORAX: The expression charts where Dotted and Checked are heterozygote are possible, but the expression chart where Tetraltera is heterozygote fails because the only possible result of crossing Dotted and Checked parents in that expression would be Tetraltera offspring, thus the expression chart fails leaving only the above mentioned two.

Results of Cross 2:

```
---------------vial-C--------------
|Parents: m vB MC i1    f vB MC i1  |
|Traits:  ANTENNAE THORAX           |
| SEX   #   ANTENNAE   THORAX       |
------------------------------------
|  f   11   MISSING    TETRALTERA   |
|  m    3   MISSING    TETRALTERA   |
|  f    8   MISSING    DOTTED       |
|  m    3   MISSING    DOTTED       |
|  f   14   MISSING    CHECKED      |
|  m    7   MISSING    CHECKED      |
|  m    4   CRINKLED   TETRALTERA   |
|  m    4   CRINKLED   DOTTED       |
|  m   11   CRINKLED   CHECKED      |
------------------------------------
```

Examining the cross data from the first cross we note that:

1. For trait ANTENNAE: Crossing a Missing male from vial B with a Missing female from vial B resulted in 19 Crinkled males, 13 Missing males, and 33 Missing females.
2. For trait THORAX: Crossing a Checked male from vial B with a Checked female from vial B resulted in 7 Tetraltera males, 11 Tetraltera females, 7 Dotted males, 8 Dotted females, 18 Checked males and 14 Checked females.

From this **Redescription** we make cross plans:

1. For trait ANTENNAE:
    a. Cross a Missing male with a Crinkled female.

   We also remove the plan to cross a Missing male from vial B with a Missing female from vial B since the cross was just done.
2. For trait THORAX:
    a. Repeat the cross of a Checked male from vial B with a Checked female from vial B.

   We also remove the plan to cross a Checked male from vial B with a

Checked female from vial B since this cross was just done.

Examining our hypotheses, we determine that:

1. For trait ANTENNAE: The expression chart where Crinkled is dominant since crossing Missing parents could never result in a Crinkled offspring. The expression chart where Missing is dominant is unlikely since the only explanation would require that a statistically unlikely (< .01) event (low offspring numbers) would have to have occurred for this expression chart to explain the data.
2. For trait THORAX: The expression chart where Dotted is heterozygous is impossible since crossing Checked parents could never result in Tetraltera or Checked offspring. The expression chart where Checked is heterozygous remains possible.

Exiting **Test and Refine Hypothesis**, since our hypothesis for ANTENNAE is no longer likely, we reach the step of **Check Alternate Hypotheses**, but since we do not have a final hypothesis for ANTENNAE yet we return to the step of **Generate Hypotheses** where:

1. For trait ANTENNAE: We decide that since Simple Dominance failed and there are some indications from the last cross, the next hypothesis we should consider is Simple Dominance and Sex Linkage. Within this hypothesis we are further able to determine that there are two possible expression charts:
   a. Missing is dominant:

Male	Female
A_ - Missing	AA - Missing
	AB - Missing
B_ - Crinkled	BB - Crinkled

   b. Crinkled is dominant:

Male	Female
A_ - Crinkled	AA - Crinkled
	AB - Crinkled
B_ - Missing	BB - Missing

Bringing these hypotheses up to date for the two crosses already done, we note that the second expression chart could not explain the second cross since crossing Missing parents could never result in Crinkled offspring, thus this expression chart fails. The other expression chart remains likely.

Since each trait now has a likely hypothesis that has only one expression chart, we do not need to reenter **Test and Refine Hypotheses**, but go straight to **Check Alternate Hypotheses**, where we would consider a number of other hypotheses (which would require a lot of space to show without really demonstrating anything new). What is determined is that other expression charts are still possible and we still need to get more data.

Returning to **Test and Refine Hypotheses** we plan to cross a Missing Checked male from vial B with a Crinkled Checked female from vial B

Results of Cross 3:

```
----------------vial-D---------------
|Parents: m vB MC il   f vB CC il   |
|Traits:  ANTENNAE THORAX           |
| SEX   #   ANTENNAE   THORAX        |
-------------------------------------
|  f   11   MISSING    TETRALTERA    |
|  f    8   MISSING    DOTTED        |
|  f   17   MISSING    CHECKED       |
|  m    5   CRINKLED   TETRALTERA    |
|  m   10   CRINKLED   DOTTED        |
|  m   19   CRINKLED   CHECKED       |
-------------------------------------
```

Examining the cross data from the first cross we note that:

1. For trait ANTENNAE: Crossing a Missing male from vial B with a Crinkled female from vial B resulted in 34 Crinkled males and 36 Missing females.
2. For trait THORAX: Crossing a Checked male from vial B with a Checked female from vial B resulted in 5 Tetraltera males, 11 Tetraltera females, 10 Dotted males, 8 Dotted females, 19 Checked males and 17 Checked females.

From this **Redescription** we make cross plans:

1. For trait ANTENNAE: We remove the plan to cross a Missing male from vial B with a Crinkled female from vial B since the cross was just done.
2. For trait THORAX:
    a. Repeat the cross of a Checked male from vial B with a Checked female from vial B.

   We also remove the plan to cross a Checked male from vial B with a Checked female from vial B since this cross was just done.

Examining our hypotheses we determine that:

1. For trait ANTENNAE: The remaining expression chart still fits the data.
2. For trait THORAX: The remaining expression chart still fits the data.

Exiting **Test and Refine Hypothesis** since we have a likely hypothesis for each trait we again reach the **Check Alternate Hypotheses**. This time the exit condition is met so we leave Experiment and enter **Check Results**.

Entering the step **Check For Modifiers**, we note that we have determined that one trait is the result of Sex Linkage and the other is not, therefore the two modifiers we would consider, Autosomal Linkage and Pleiotropy are not possible, thus we skip straight to the Review-Solution step, where we do a statistical analysis of our solution (not included).

APPENDIX H

An Example of an Ideal Justification in a Tutoring Session

A student's ability to explain his or her solution is an important feature of model-based problem solving. In this, we present an example of an explanation for a problem of moderate difficulty. The example serves as an illustration of how we define and operationalize understanding and what a student should be able to do. The explanation is structured around our **Strategy Tree**. At the end of each section, we have included questions that could be asked to elicit a student's explanations. While not all of the answers to these questions will be able to be easily translated into the MENDEL system, they do provide a basis for deciding which procedures could be routinized so that a computer could prompt students to provide explanations that could be evaluated.

A student wouldn't be asked to explain every step of a problem solution or provide the same type of explanation at every stage of their problem-solving experience. It may be that as their experience increases with a variety of problem types the less they would be asked to provide explanations in terms of meiosis. Once students were judged to be proficient at explaining solutions to a set of problems in terms of meiosis, their knowledge of meiosis would then only be reevaluated when they encounter new inheritance patterns or modifiers. In addition to explanation in terms of meiosis, there are other explanations (at a non-causal model level) that will involve providing definitions, empirical associations and manipulating allelic symbols. What follows is a solution to one problem that includes explanations, definitions, empirical associations, and meiosis. The problem itself could include: simple dominance, codominance, sex linkage, and autosomal linkage. The problem begins with a parental vial of field collected organisms.

Parental Vial:

```
---------------vial-A--------------
|Parents: Parental Vial             |
|Traits:  EYES WINGS                |
| SEX   #  EYES        WINGS        |
-----------------------------------
|  f    5  APRICOT     QUILTED      |
|  m    7  APRICOT     QUILTED      |
|  f    3  YELLOW      SHORT        |
|  m    2  YELLOW      SHORT        |
|  f    4  YELLOW      QUILTED      |
|  m    2  YELLOW      QUILTED      |
|  f    3  GARNET      SHORT        |
|  f    1  GARNET      QUILTED      |
|  m    3  GARNET      QUILTED      |
|  f    6  APRICOT     SHORT        |
-----------------------------------
```

Redescription: In Vial-A, it should be noted that there are 2 traits (EYES and WINGS); that EYES has 3 variations (Apricot, Yellow, and Garnet) and WINGS has 2 variations (Quilted and Short). It can also be noted that there are 2 missing phenotype classes Garnet Short males and Apricot Short males (missing phenotypes by sex classes may indicate sex linkage).

Questions to Elicit a Student's Explanations:

1. How many traits are there, how many variations per trait?
2. Why is it important to note missing phenotype classes?

Entertain Initial Hypothesis: On the basis of this redescription, it is possible to entertain 2 hypotheses:

1. that EYES is the result of codominance, and;
2. that WINGS is the result of simple dominance.

To explain at this point is to indicate that by definition codominance is one situation in which there are 3 variations for each trait, and similarly simple dominance is characterized by 2 variations for a trait. (By establishing these initial hypotheses there is additional "baggage" concerning the number of genotypes and how genotypes map to phenotypes, more on this under **Explain cross**). There are other hypotheses consistent with the number of variations exhibited for each trait but they are more complex than codominance and simple dominance, and therefore can be ignored until the simple hypotheses are shown to be inadequate.

Questions to Elicit a Student's Explanations:

1. Why do you think that the inheritance pattern for EYES is codominance?
2. Why do you think that the inheritance pattern for WINGS is simple dominance?
3. Why are you entertaining an hypothesis at this time?
4. What do you mean by codominance, by simple dominance?

Perform A Cross: Even though there are some crosses that could produce more knowledge and therefore be more efficient, it is more important that a student can explain any cross results. The first cross was a cross between parents with unlike variations for each trait. This happens to be a very efficient first cross.

Results of Cross 1:

```
----------------vial-B---------------
|Parents: m vA YS i1   f vA AQ i1   |
|Traits:  EYES WINGS                |
| SEX    #   EYES        WINGS      |
-------------------------------------
|  f     7   APRICOT    SHORT       |
|  m    10   APRICOT    QUILTED     |
|  f    13   YELLOW     SHORT       |
|  m     8   YELLOW     QUILTED     |
-------------------------------------
```

Questions to Elicit a Student's Explanations:

1. Why did you do this particular cross?
2. Could have performed other crosses that would have been equally useful?
3. What makes a cross useful?

Redescribe cross results: Two of the three variations for the EYES trait appear and both of the variations for the WINGS trait appeared. There are no new variations for either trait. This is important since if new variations for either trait had appeared it would have meant that our initial hypotheses were wrong.

Questions to Elicit a Student's Explanations:

1. Why was it important to note that there are no new variations for either trait?
2. Why is it important to redescribe the data after each new cross?

Explain Cross Results: It is possible to solve for both traits at the same time although it is acceptable to solve for one trait at a time. Since the EYES trait was assumed to be codominant, it is possible, using these cross results, to assume that either Apricot or Yellow is the heterozygous condition and the other is a homozygote. This assumption can be explained by invoking a codominance **Expression Chart**. The function of the chart is to initiate genotypic thinking, in addition to the phenotypic level of traits and variations. An **Expression Chart** for codominance is:

Genotype:	Phenotype:
AA	Variation 1
AB	Variation 2
BB	Variation 3

In addition, in order to fully explain these results at the level of allelic symbols, it is necessary to have general information about what it means to do a cross, including that:

1. in codominance or simple dominance each variation is represented by a pair of letters;
2. this pair of letters is conserved from parents to offspring;
3. each parent donates one of their letters for a given trait to each of their offspring;
4. a **Punnett Square** can be used to represent cross results. For the above genotypes (assuming codominance) using a **Punnett Square** it is known that:
    1. AA x AA - AA (Variation 1)
    2. BB x BB - BB (Variation 3)
    3. AA x BB - AB (Variation 2)
    4. AB x AB - AA (Variation 1); AB (Variation 2); BB (Variation 3)
    5. AB x AA - AA (Variation 1); AB (Variation 2)
    6. AB x BB - AB (Variation 2); BB (Variation 3)

Therefore the results of Cross 1 can be explained using Cross #5 or Cross #6.

The second trait, assuming simple dominance, can be explained if one parent was heterozygous and the second was homozygous recessive. This explanation requires that a student understand dominance (that anytime one allele, called the dominant, is present the individual will have a particular variation) and recessive (that the recessive variation will not be exhibited if there is a dominant allele combined with it, the recessive variation will only be expressed in the homozygous condition). Understanding that:

1. homozygous means two of the same letters;
2. it is possible to be homozygous dominant or homozygous recessive;
3. heterozygous means having one of each letter; and,
4. in simple dominance it is only possible to be heterozygous dominant,

facilitates explanation. As was the case with codominance there is an

Expression Chart for Simple Dominance. It is as follows:

Genotype:	Phenotype:
AA	Variation 1
Aa	Variation 1
aa	Variation 2

Likewise, there are **Cross Equations** that can be manipulated in a Punnett Square that serve the same function as those for codominance. They are:

1. AA x AA - AA (Variation 1)
2. aa x aa - aa (Variation 2)
3. AA x Aa - AA (Variation 1); Aa (Variation 1)
4. AA x aa - Aa (variation 1)
5. Aa x Aa - AA (Variation 1); Aa (Variation 1);
   aa (Variation 2)
6. Aa x aa - Aa (Variation 1); aa (Variation 2)

The **Cross Equation** that could be used to explain the results of Cross 1 is #6. At this point no more can be said about the two traits. The result of this explanation is that it has reduced the search space.

To this point the explanation of the cross has involved accurate definitions of concepts and the explanation of the cross results at the level of symbols -- this is what Mendel did, so it should not be taken lightly. However, it is possible to provide a more causal explanation in terms of meiosis and fertilization. What this does is extend the discussion about: the separation of symbols (segregation) discussed above; the fact that we considered the two traits separately (independent assortment); and, the use of the **Punnett Square** from the level of the abstract, non-causal to a consideration of segregation and independent assortment at the level of genes and chromosomes (Meiosis) and fertilization (represented by a Punnett square). Diagrams are an appropriate way to make explanations graphically explicit at this level. See the next two Figures.

An Abstract Representation of Meiosis for a Monohybrid Cross

```
                                 -----------------
Abstract Cell with two___________|   ---   ---   |
Chromosomes (A a)_______________>|   |A|   |a|   |
                                 |   | |   | |   |
                                 |   | |   | |   |
                                 |   | |   | |   |
                                 |   | |   | |   |
                                 |   ---   ---   |
                                 -----------------
                           -----------------------------
Chromosomes Replicate_____>|   --- ---       --- ---   |
                           |   |A| |A|       |a| |a|   |
                           |   | | | |       | | | |   |
                           |   | |*| |       | |*| |   |
                           |   | | | |       | | | |   |
                           |   | | | |       | | | |   |
                           |   --- ---       --- ---   |
                           -----------------------------
                     -----------------       -----------------
Cell Division 1 ____>|    --- ---    |       |    --- ---    |
                     |    |A| |A|    |       |    |a| |a|    |
                     |    | | | |    |       |    | | | |    |
                     |    | |*| |    |       |    | |*| |    |
                     |    | | | |    |       |    | | | |    |
                     |    | | | |    |       |    | | | |    |
                     |    --- ---    |       |    --- ---    |
                     -----------------       -----------------
            -----------     -----------     -----------     -----------
Cell      _ |   ---   |     |   ---   |     |   ---   |     |   ---   |
Division 2 >|   |A|   |     |   |A|   |     |   |a|   |     |   |a|   |
            |   | |   |     |   | |   |     |   | |   |     |   | |   |
            |   | |   |     |   | |   |     |   | |   |     |   | |   |
Gametes ___>|   | |   |     |   | |   |     |   | |   |     |   | |   |
            |   | |   |     |   | |   |     |   | |   |     |   | |   |
            |   ---   |     |   ---   |     |   ---   |     |   ---   |
            -----------     -----------     -----------     -----------
```

An Abstract Representation of Meiosis for a Monohybrid showing Sex Linkage (in male only). What happens in female is the same as the Monohybrid but with no Linkage representation.

```
                                 -----------------
                              ___|__ ---   --- __|___
                 X Chromosome    |  >|A|   | |<  |    Y Chromosome
                                 |   | |   | |   |
                                 |   | |   ---   |
                                 |   | |         |
                                 |   | |         |
                                 |   ---         |
                                 -----------------
                      ____ -----------------------------
Chromosomes Replicate     >|   --- ---       --- ---   |
                           |   |A| |A|       | | | |   |
                           |   | | | |       | | | |   |
                           |   | |*| |       ---*---   |
                           |   | | | |                 |
                           |   | | | |                 |
                           |   --- ---                 |
                           -----------------------------
                ____ -----------------       -----------------
Cell Division 1     >|    --- ---    |       |   --- ---     |
                     |    |A| |A|    |       |   | | | |     |
                     |    | | | |    |       |   | | | |     |
                     |    | |*| |    |       |   ---*---     |
                     |    | | | |    |       |               |
                     |    | | | |    |       |               |
                     |    --- ---    |       |               |
                     -----------------       -----------------
            -----------     -----------     -----------     -----------
Cell      _ |   ---   |     |   ---   |     |   ---   |     |   ---   |
Division 2 >|   |A|   |     |   |A|   |     |   | |   |     |   | |   |
            |   | |   |     |   | |   |     |   | |   |     |   | |   |
        ___ |   | |   |     |   | |   |     |   ---   |     |   ---   |
Gametes    >|   | |   |     |   | |   |     |         |     |         |
            |   | |   |     |   | |   |     |         |     |         |
            |   ---   |     |   ---   |     |         |     |         |
            -----------     -----------     -----------     -----------
```

Questions to Elicit a Student's Explanations:

1. What do you mean when you use the terms genotype, phenotype, heterozygous, and homozygous?
2. What do the letters that you are using represent?
3. How do you know which letters the offspring could have?
4. When you are working with both traits at the same time, why isn't it possible to have both A's from one parent around the outside of your **Punnett Square**?
5. What value is there in using a **Punnett Square**? Does it represent anything that goes on in the real world of organisms?
6. Could you explain what gametes have to do with your solution? What do you mean by gamete? Is it possible for you to point to anything in your solution that represents a gamete?
7. What do you mean by the terms gene and allele? Are genes and alleles related in any way? How?
8. Are chromosomes at all related to your solution? How?
9. Is there any way of knowing how many chromosomes the organisms in this problem have? How many of their total chromosomes do you think are involved in this problem?
10. Is it possible for you to draw a picture of how you imagine the chromosomes in this problem look?
11. Could you put the letters that you have been using in your problem solution on your chromosome drawings?
12. Can you demonstrate how these diagrams, with the letters on them go form the parents that you chose for your cross to their offspring?
13. Could you explain what meiosis has to do with your solution?
14. Suppose your were told that other students drew the following chromosome/gene models that differ from yours, what do you think they do not know about genetics? If you were a teacher how would you try to straighten out their understanding?
15. Could you explain how meiosis is related to your problem solution?
16. What are the different genotypes that can exist in simple dominance, in codominance?
17. How many phenotypes are possible in simple dominance, in codominance? How can these different phenotypes be matched to the genotypes that you have just described?
18. Can you explain, using the letters that you have been using, how the two EYES variations that you crossed could have produced the resulting offspring? Can you do the same thing for the WINGS trait? Are there any other possible genotype that the parents could have had that would have produced the same offspring phenotypes?

Plan Cross: It is still true that any cross is a good cross as long as it can be explained, even though it makes more sense from the standpoints of efficiency and emulating what a geneticist does to take advantage of information that has been inferred, from the results of previous crosses, about the genotypes of individuals or classes of individuals. It is in Cross **Planning** where a solver takes advantage of gained information. When asked to explain why Cross 2 was done, a mature solver would respond that: given the **Cross Equations** for codominance, it makes sense to cross individuals of the same phenotype from the offspring Vial B since it is possible to **predict** that crosses among likes would give only offspring with the same phenotype as the parents and thus are likely homozygotes. If this is the case then the other variation in Vial B is likely the heterozygote and the variation missing from Vial B is the second homozygote. If the cross produces offspring with all three variations, then the parental variation is the heterozygote and the other two variations are homozygotes. With this information, it is easy to assess the results of the second cross and to plan additional confirmatory crosses. The WINGS trait will be considered later. The results of the second cross are as follows:

Results of Cross 2:

```
----------------vial-C--------------
|Parents: m vB AQ il    f vB AS il   |
|Traits:  EYES WINGS                 |
| SEX   #   EYES         WINGS       |
-------------------------------------
|  f    4   APRICOT      SHORT       |
|  m    4   APRICOT      SHORT       |
|  f    1   YELLOW       SHORT       |
|  m    3   YELLOW       SHORT       |
|  f    2   GARNET       SHORT       |
|  m    3   GARNET       SHORT       |
|  f    2   APRICOT      QUILTED     |
|  m    5   APRICOT      QUILTED     |
|  f    3   YELLOW       QUILTED     |
|  m    1   YELLOW       QUILTED     |
|  f    1   GARNET       QUILTED     |
|  m    3   GARNET       QUILTED     |
-------------------------------------
```

Questions to Elicit a Student's Understanding:

Redescribe Cross Results: When an Apricot Short individual was crossed with an Apricot Quilted individual (both from Vial B), all three EYES variations and both WINGS variations resulted. There are no missing phenotype by sex classes.

Explain Cross Results: Given these cross results, it is possible to be confident that the EYES trait is codominant and that the apricot variation is the heterozygous, and that the yellow, and garnet variations are homozygous. The reason for this is that two apricots when crossed produced offspring with all three variations. This can be explained with the abstract **Cross Equation** AB x AB - AA, AB, and BB. The AB genotype is the class of apricot-EYEd organisms. Nothing new is known about the second trait.

Plan Cross: It is now important to plan a cross that will provide additional information about the WINGS trait. Therefore a cross between two like individuals from Vial C is useful. The utility of this cross lies in the interpretation of **Cross Equation** #2. A cross between unlike parents that results in offspring with the variations of each parent means that the offspring with one variation are homozygous recessive and the offspring with the second variation are heterozygous. Thus, the third cross is either **Cross Equation** #2 or #5 (from above). The following cross was done to satisfy the **Cross Planning**.

Results of Cross 3:

```
----------------vial-D--------------
|Parents: m vC GS il   f vC GS il   |
|Traits:  EYES WINGS                |
| SEX   #   EYES        WINGS       |
-------------------------------------
|  f   13   GARNET      SHORT       |
|  m    2   GARNET      SHORT       |
|  m    5   GARNET      QUILTED     |
-------------------------------------
```

Redescribe Cross Results: Two Garnet-EYEd individuals produced only Garnet-eyed individuals and two Short-WINGed individuals produced offspring with both short and quilted EYES. Note that there are no Garnet Quilted females. This is noted because of a possible connection to the **Redescription** of the initial vial in which sex-linkage was noted as something to watch for.

Explain Cross Results: The results of the cross of the two Garnet individuals is consistent with the information that all Garnets's must be homozygous. For the EYES, trait since two like individuals were crossed and the offspring were of two variations this is a good indication (given simple dominance) that the two short individuals were heterozygous (therefore short is due to a dominant allele). The results can be explained by the **Cross Equation** Aa x Aa = Aa and aa. Since a sex class with a missing variation is a good indicator of sex linkage it is reasonable to see if these results could be explained by some sex link cross. The first task is to decide on which (or both of) the traits might be sex linked. It is relatively easy to be sure that the EYES trait is not sex linked as both male and female offspring of each variation exist. In order to explain this and to begin to explore the possibility that the WINGS trait might be sex linked it is useful to consider a symbolic **Expression Chart** for sex linkage. Such an **Expression Chart** is illustrated below and presumes that the solver's knows that:

1. sex linkage implies that the allele letters are linked to the X chromosome;
2. the chromosomal make up of females is XX and that of males is XY;
3. the Y chromosome contains little active genetic material;
4. Therefore a single dose of an allele causes the expression of the appropriate phenotype in males. There are no heterozygous males in sex linkage.

Genotype:	Phenotype:	
	Simple Dominance	Codominance
$X^A X^A$	(F) Variation 1	(F) Variation 1
$X^A X^a$	(F) Variation 1	(F) Variation 2
$X^a X^a$	(F) Variation 2	(F) Variation 3
$X^A Y$	(M) Variation 1	(M) Variation 1
$X^a Y$	(M) Variation 2	(M) Variation 3

Cross Equations for sex linkage will also be part of an explanation. These rules which can be used to explain or to predict are:

1. $X^A X^A$ x $X^A Y$ - $X^A X^A$ and $X^A Y$ (uninteresting)
2. $X^a X^a$ x $X^a Y$ - $X^a X^a$ and $X^a Y$ (uninteresting)
3. $X^A X^A$ x $X^a Y$ - $X^A X^a$ and $X^A Y$ (uninteresting)
4. $X^A X^a$ x $X^a Y$ - $X^A X^a$; $X^a X^a$; and $X^A Y$; $X^a Y$ (uninteresting)
5. $X^a X^a$ x $X^A Y$ - $X^A X^a$ and $X^a Y$ (interesting, one variation for each sex)
6. $X^A X^a$ x $X^A Y$ - $X^A X^A$; $X^A X^a$; and $X^A Y$; $X^a Y$ (interesting, since only one variation for the female)

The reason that some crosses are uninteresting is that they cannot be used to distinguish between a sex linkage and a non-sex linkage situation. Now the task is to explain the results of crossing two like parents where the offspring have both variations, and where there are no females of the variation not exhibited in the parents. Producing offspring with two variations can be explained if the short variation is due to the dominant allele and the two parents were heterozygous (Simple dominance **Cross Equation** #5). There is also a strong possibility that sex linkage is involved as the missing female variation can be explained using sex linkage **Cross Equation** #6. A solution is close at hand, yet another cross to further explore the tentative solution just described is warranted.

Plan Cross: A cross that helps to confirm that the EYES trait is codominant and that the WINGS trait is simple dominant (with short due to a dominant allele) would be to cross a recessive female (any Quilted individual) with a Short male (if sex linkage is operating then any dominant male will only have a single dominant allele). See the expression chart above for clarification and see Sex Linkage **Cross Equation** #5 for the cross that is being planed to test for sex linkage. If sex linkage is involved there will be short females and quilted males, and nothing else in the offspring. The cross that results is as follows:

Results of Cross 4:

```
----------------vial-E---------------
|Parents: m vD GS il   f vA GQ il   |
|Traits:  EYES WINGS                     |
| SEX   #   EYES        WINGS            |
-----------------------------------------
|  m    8   GARNET      QUILTED          |
|  f   14   GARNET      SHORT            |
-----------------------------------------
```

Redescribe Cross Results: There were Garnet Quilted females and Garnet Short males.

Explain Cross Results: The results are consistent with the codominance hypothesis for the EYES trait -- Garnet was assumed to be one of the two homozygotes therefore only Garnet offspring would be expected. The hypothesis of sex linkage receives additional support since the prediction made in the **Cross Plan** was confirmed. One additional cross might be done to further confirm the hypotheses.

Plan Cross: If Apricot is the heterozygous variation then a cross between two apricot parents should result in offspring with all three EYES variations. A check of the sex linkage hypothesis would be to repeat **Cross Equation #5** using different individuals. If sex linkage is involved it would be expected (for the same reasons given above) that there would only be recessive males (quilted) and dominant females (short). The results of the cross follow:

Results of Cross 5:

```
---------------vial-F---------------
|Parents: m vC AS il   f vA AQ il   |
|Traits:  EYES WINGS                |
| SEX   #   EYES       WINGS        |
------------------------------------
|  m    7   YELLOW     QUILTED      |
|  f   10   APRICOT    SHORT        |
|  f    3   YELLOW     SHORT        |
|  m    5   GARNET     QUILTED      |
|  f    4   GARNET     SHORT        |
|  f    4   APRICOT    QUILTED      |
------------------------------------
```

Explain Cross Results: This cross was done as a check on the last one. The results were the same -- all of the males were quilted and all of the females were short. This can be explained in the same way as was the results of cross 4. At this point the problem can be considered solved. It should be noted that it is normally a good idea to check for autosomal linkage at the end of a problem if it hasn't been considered along the way. In this particular problems there is no need to check if the solver knows that autosomal linkage means that the genes for two traits are located on the same pair of autosomal chromosomes. (However, if such a check were to be done, the last four Figures show the appropriate abstract meiotic mechanisms). Since it has been assumed that there were only two pair of genes in this problem, one controlling the EYES trait and the other the WINGS trait, and since it was established that the genes for the WINGS trait were linked to the X chromosome and that the genes for the EYES trait could not be on the X chromosome there is no reason to pursue the hypothesis of autosomal linkage.

It has been assumed throughout that the solver understands that:

1 symbols disjoin so that only one goes to each gamete (sperm or egg).
2. The causal mechanism underlying the law of segregation is meiosis.

To explain this involves the following information:

1. Letters represent genes/alleles
2. Each cell in the parent has two letters (genes/alleles)
3. Different combinations of letters can produce different phenotypes (the expression charts).
4. Gene/allele are on chromosomes
5. Chromosomes come in pairs (homologous pairs).
6. One member of each gene pair is on each homologous chromosome.
7. Meiosis can be graphically represented as in Figures the last four figures.

An Abstract Representation of Meiosis for a Dihybrid Problem in which there is Incomplete Autosomal Linkage.

```
                                       -----------------
  Abstrace Cell With Four             |   ---   ---     |
  Chromosomes (A a B b) _____________>|   |A|   |a|     |
                                      |   | |   | |     |
                                      |   | |   | |     |
                                      |   | |   | |     |
                                      |   |B|   |b|     |
                                      |   ---   ---     |
                                       -----------------

                             ------------------------------
 Chromosomes Replicate  ____>|   --- ---       --- ---     |
                             |   |A| |A|       |a| |a|     |
                             |   | | | |       | | | |     |
                             |   | |*| |       | |*| |     |
                             |   | | | |       | | | |     |
                             |   |B| |B|       |b| |b|     |
                             |   --- ---       --- ---     |
                             ------------------------------

                            ------------------------------
 Same as above but after   |   --- ---       --- ---     |
 a cross-over event ______>|   |A| |A|       |a| |a|     |
                           |   | | | |       | | | |     |
                           |   | |*| |       | |*| |     |
                           |   | | | | Cross | | | |     |
                           |   |B| |b|<_____>|B| |b|     |
                           |   --- ---  Over --- ---     |
                            ------------------------------

                       ----------------        ----------------
 Cell Division 1 ____>|   --- ---      |      |    --- ---     |
                      |   |A| |A|      |      |    |a| |a|     |
                      |   | | | |      |      |    | | | |     |
                      |   | |*| |      |      |    | |*| |     |
                      |   | | | |      |      |    | | | |     |
                      |   |B| |b|      |      |    |B| |b|     |
                      |   --- ---      |      |    --- ---     |
                       ----------------        ----------------

             ----------    ----------    ----------    ----------
 Cell       |   ---    |  |   ---    |  |   ---    |  |   ---    |
 Division 2>|   |A|    |  |   |A|    |  |   |a|    |  |   |a|    |
            |   | |    |  |   | |    |  |   | |    |  |   | |    |
 Gametes __>|   | |    |  |   | |    |  |   | |    |  |   | |    |
            |   | |    |  |   | |    |  |   | |    |  |   | |    |
            |   |B|    |  |   |b|    |  |   |B|    |  |   |b|    |
            |   ---    |  |   ---    |  |   ---    |  |   ---    |
             ----------    ----------    ----------    ----------
```

An Abstract Representation of Meiosis for a Dihybrid Problem in which there is Complete Autosomal Linkage.

```
                                 ----------------
Abstract Cell With Four_________|    ---    ---   |
Chromosomes (A a B b)          >|    |A|    |a|   |
                                |    | |    | |   |
                                |    | |    | |   |
                                |    | |    | |   |
                                |    |b|    |B|   |
                                |    ---    ---   |
                                 ----------------

                            ----------------------------
Chromosomes Replicate  ___>|   --- ---      --- ---    |
                           |   |A| |A|      |a| |a|    |
                           |   | | | |      | | | |    |
                           |   | |*| |      | |*| |    |
                           |   | | | |      | | | |    |
                           |   |b| |b|      |B| |B|    |
                           |   --- ---      --- ---    |
                            ----------------------------

                        ----------------        ----------------
Cell Division 1 ____>|   --- ---      |      |    --- ---     |
                        |   |A| |A|      |      |    |a| |a|     |
                        |   | | | |      |      |    | | | |     |
                        |   | |*| |      |      |    | |*| |     |
                        |   | | | |      |      |    | | | |     |
                        |   |b| |b|      |      |    |B| |B|     |
                        |   --- ---      |      |    --- ---     |
                        ----------------        ----------------

                  ----------    ----------    ----------    ----------
Cell             _|   ---    |  |   ---    |  |   ---    |  |   ---    |
Division 2 >|   |A|    |  |   |A|    |  |   |a|    |  |   |a|    |
                 |   | |    |  |   | |    |  |   | |    |  |   | |    |
                 |   | |    |  |   | |    |  |   | |    |  |   | |    |
Gametes   ___>|   | |    |  |   | |    |  |   | |    |  |   | |    |
                 |   |b|    |  |   |b|    |  |   |B|    |  |   |B|    |
                 |   ---    |  |   ---    |  |   ---    |  |   ---    |
                  ----------    ----------    ----------    ----------
```

An Abstract Representation of Meiosis for a Dihybrid problem in which there is no Linkage.

```
                          ------------------------------
Abstrace Cell With Four__ |   ---   ---     ---   ---   |
Chromosomes (A a B b)   >|   |A|   |a|     |B|   |b|   |
                          |   | |   | |     | |   | |   |
                          |   | |   | |     | |   | |   |
                          |   | |   | |     ---   ---   |
                          |   | |   | |                 |
                          |   ---   ---                 |
                          ------------------------------
                           ----------------------------
                           |   --- ---       --- ---  |
Chromosomes Replicate ____>|   |A| |A|       |a| |a|  |
A's and B's could be       |   | | | |       | | | |  |
aligned differently        |   | |*| |       | |*| |  |
since the process of       |   | | | |       | | | |  |
alignment is a random      |   | | | |       | | | |  |
one.  The alternate        |   --- ---       --- ---  |
alignment would provide    |   --- ---       --- ---  |
2 additional classed of    |   |b| |b|       |B| |B|  |
gametes.                   |   | |*| |       | |*| |  |
                           |   | | | |       | | | |  |
                           |   --- ---       --- ---  |
                           ----------------------------
                     ----------------        ----------------
Cell Division 1 ____>|   --- ---    |        |   --- ---    |
                     |   |A| |A|    |        |   |a| |a|    |
                     |   | | | |    |        |   | | | |    |
                     |   | |*| |    |        |   | |*| |    |
                     |   | | | |    |        |   | | | |    |
                     |   | | | |    |        |   | | | |    |
                     |   --- ---    |        |   --- ---    |
                     |   --- ---    |        |   --- ---    |
                     |   |b| |b|    |        |   |B| |B|    |
                     |   | |*| |    |        |   | |*| |    |
                     |   | | | |    |        |   | | | |    |
                     |   --- ---    |        |   --- ---    |
                     ----------------        ----------------
```

```
              -----------       -----------       -----------       -----------
Cell         |    ---    |     |    ---    |     |    ---    |     |    ---    |
Division 2 ->|    |A|    |     |    |A|    |     |    |a|    |     |    |a|    |
             |    | |    |     |    | |    |     |    | |    |     |    | |    |
             |    | |    |     |    | |    |     |    | |    |     |    | |    |
             |    | |    |     |    | |    |     |    | |    |     |    | |    |
Gametes    ->|    | |    |     |    | |    |     |    | |    |     |    | |    |
             |    ---    |     |    ---    |     |    ---    |     |    ---    |
             |    ---    |     |    ---    |     |    ---    |     |    ---    |
             |    |b|    |     |    |b|    |     |    |B|    |     |    |B|    |
             |    | |    |     |    | |    |     |    | |    |     |    | |    |
             |    | |    |     |    | |    |     |    | |    |     |    | |    |
             |    ---    |     |    ---    |     |    ---    |     |    ---    |
              -----------       -----------       -----------       -----------
```

END

U.S. Dept. of Education

Office of Education
Research and
Improvement (OERI)

ERIC

Date Filmed

March 29, 1991